



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

상이한 사용자 그룹이 VUI 이용 시 경험하는 어려움에 대한 이해

사용 빈도와 최신성을 바탕으로 한
사용자 구분 기준에 근거하여

2020 년 8 월

서울대학교 융합과학기술대학원

융합과학부 디지털정보융합전공

정 현 훈

상이한 사용자 그룹이 VUI 이용 시 경험하는 어려움에 대한 이해

사용 빈도와 최신성을 바탕으로 한

사용자 구분 기준에 근거하여

지도 교수 서 봉 원

이 논문을 공학석사 학위논문으로 제출함

2020 년 7 월

서울대학교 융합과학기술대학원

융합과학부 디지털정보융합전공

정 현 훈

정현훈의 공학석사 학위논문을 인준함

2020 년 7 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

초 록

음성 사용자 인터페이스 (Voice User Interface, VUI)가 확산되며, VUI는 이제 기술 수용에 적극적인 사용자층만을 위한 인터페이스가 아닌 다수가 일상적으로 사용할 수 있는 인터페이스가 되었다. 기술 수용 및 활용 적극성 관점에서 사용자를 분류하여 이해할 필요가 있음에도 불구하고, 현재 HCI 커뮤니티의 VUI에 대한 사용자 연구는, 이와 같은 비교 연구를 진행하고 있는 경우는 드물다. 본 연구는, 이용빈도와 최신성에 입각하여 Heavy User와 Light User로 사용자 그룹을 분리해, 이들이 해당 인터페이스를 사용하며 겪는 어려움에 초점을 맞추어 알아본다. 각각의 사용자가 어떤 어려움을 겪는지, 그리고 이 어려움에 어떻게 반응하거나 대응하는지에 대해 알아본다. 이 결과를 통해 VUI의 디자인과 개발 측면에서의 개선점이 무엇이 있을지 도출하고자 한다. 본 연구는, 네이버 클로바 AI (Clova AI, Naver Corporation) 연구 조직에서 디자인하고 개발한 식당 예약 음성 에이전트, AiCall이라는 제품을 활용하여 유저 스터디를 진행하였다. 필자는 연구의 결과로, Heavy User와 Light User가 보고한 어려움의 종류와 순위에 차이가 있음을 발견했다. 또한, 각 사용자 그룹이 어려움에 반응하거나 대응하는 방식의 경향성의 차이에 대해서도 알 수 있었다. Heavy User는 문제 상황에 대해 적극적이었고 직설적이었으며, Light User는 문제 상황에 대해서 수동적인 태도를 보였다. 본 연구는 기술 수용 및 활용 적극성의 기준에 따라 분류된 각 사용자 그룹이 VUI 이용 시 겪는 어려움에 대해 이해하는 것을 통해, 이러한 어려움을 개선하기 위해 VUI의 개발 및 제작 방향은 어떠해야 하는지 제안한다.

주요어 : 음성 사용자 인터페이스(Voice User Interface, VUI), 대화형 사용자 인터페이스 (Conversational User Interface, CUI), 음성 사용자 경험 (Voice UX), 대화형 사용자 경험 (Conversational UX)

학 번 : 2017 - 29610

목 차

1. 서 론 (Introduction)	1
1-1. 연구의 배경 (Background).....	1
1-2. 연구의 내용 (Research Goal).....	3
1-2-1. 사용자 그룹 구분	3
1-2-2. 연구 질문 (Research Question).....	5
2. 관련 연구 (Related Work).....	7
2-1. 소통 실패와 복귀 전략.....	7
2-1-1. VUI 상에서의 소통 실패와 복귀 전략	7
2-1-2. 챗 인터페이스 상에서의 소통 실패와 복귀 전략	10
2-2. Heavy User와 Light User 사용자 그룹 비교 연구	12
2-3. VUI 적용 기술 및 제품 그리고 서비스의 발전 흐름	13
2-3-1. 대화식 음성 응답	13
2-3-2. 스마트 디바이스의 VUI.....	15
2-3-3. AiCall의 기술 및 적용 영역 상의 위치	16
3. 연구 방법 (Methodology)	18
3-1. AiCall	18
3-1-1. AiCall 시스템의 구성	18
3-1-2. AiCall 대화 디자인	20
3-1-3. AiCall 에이전트 페르소나 디자인	23
3-1-4. AiCall 데이터 구축 및 자연어 이해	25
3-1-5. AiCall 전화 환경 음성 인식.....	25
3-2. User Study	26
3-2-1. 참여자 (Participants).....	26
3-2-2. 진행 방식 (Procedure)	28

3-3. 분석 방법 (Analysis Method).....	33
3-3-1. VUI 사용 시 겪는 어려움을 이해하기 위한 분석	34
3-3-2. VUI 사용 시 겪는 어려움에 대한 대응 방식 분석	36
4. 연구 결과 (Result)	38
4-1. SUIQ-MR 결과 분석.....	38
4-2. VUI 이용 시 겪는 어려움의 유형 (Difficulty Types).....	40
4-2-1. Heavy User 그룹의 어려움 유형	44
4-2-2. Light User 그룹의 어려움 유형	45
4-3. VUI 구성 단위로 재범주화된 어려움 유형	46
4-4. VUI 사용 시 맞이하는 어려움을 대응하는 방식.....	48
4-4-1. 어려움 유형 D1에 대한 대응 방식.....	49
4-4-2. 어려움 유형 D9에 대한 대응 방식.....	55
4-4-3. 어려움 유형 D6에 대한 대응 방식.....	57
5. 논 의 (Discussion)	61
5-1. VUI 이용 행태로 파악한 각 사용자 그룹의 경향성.....	61
5-1-1. Heavy User는 적극적이며, 의사 표현이 직설적임	61
5-1-2. Light User는 수동적이며, 가이드를 필요로 함.....	64
5-2. 다양한 사용자 그룹을 고려한 VUI 개선 방향성.....	66
5-2-1. VUI 개선 우선순위에 대한 이해의 필요성	66
5-2-2. 소통 실패를 해결하기 위한 실질적 방안	68
6. 연구의 한계 (Limitations)	73
7. 결 론 (Conclusion)	75

참고문헌.....	77
-----------	----

Abstract	84
----------------	----

표 목차

[표 1].....	26
[표 2].....	27
[표 3].....	28
[표 4].....	29
[표 5].....	30
[표 6].....	31
[표 7].....	33
[표 8].....	36
[표 9].....	38
[표 10].....	39
[표 11].....	41
[표 12].....	47
[표 13].....	48
[표 14].....	49
[표 15].....	50
[표 16].....	52
[표 17].....	53
[표 18].....	54
[표 19].....	56
[표 20].....	58
[표 21].....	59
[표 22].....	70
[표 23].....	71

그림 목차

[그림 1]	19
[그림 2]	23
[그림 3]	43

1. 서 론 (Introduction)

1-1. 연구의 배경 (Background)

음성 사용자 인터페이스 (Voice User Interface, VUI)가 사람들의 생활 속으로 깊이 스며들고 있다. 과거에는 전통적인 대화식 음성 응답 (Interactive Voice Response, IVR)을 통해 제한적으로 VUI를 경험할 수 있었던 반면, 현재는 다양한 기기 혹은 서비스를 활용하여 VUI를 경험할 수 있게 되었다. Siri (Apple), Bixby (Samsung)와 같이 스마트폰에서 만날 수 있는 음성 에이전트 (Voice Agent)부터, Alexa (Amazon), Google Assistant (Google)와 같은 스마트 스피커 (Smart speaker)를 중심으로 다른 연결 기기를 통해 만나볼 수 있는 음성 에이전트까지 최근 사람들이 접할 수 있는 VUI는 다양해지고 있다.

현재의 인공 지능 기술의 발전은 VUI의 발전과 확산에 긍정적인 영향을 주고 있다. 특히, 음성 인식 (Automatic Speech Recognition, ASR), 자연어 이해 (Natural Language Understanding, NLU) 및 처리 (Natural Language Processing, NLP) 그리고 음성 합성 (Speech Synthesis) 기술의 발전은 VUI 발전 및 확산에 주요하게 기여한다. 음성 인식 기술은 사람들의 목소리를 기기 혹은 서비스가 더 잘 인식할 수 있게 만들고, 자연어 이해 기술을 중심으로 한 자연어 처리 기술은 인식된 자연어 발화를 더 잘 이해하고 처리할 수 있도록 하며, 음성 합성 기술은 처리된 내용을 바탕으로 음성 에이전트가 더욱 사람처럼 자연스럽게 발화할 수 있도록 만든다.

이러한 기술의 발전과 다양한 기기 및 서비스를 통한 VUI의 확산은, VUI에 대한 사람들의 수용도에도 영향을 미치고 있다. 2018년 VUI 및 대화형 사용자 인터페이스 (Conversational User Interface, CUI) 관련 산업 소식을 다루는 매체에서는, VUI의 시장 확산과 사용자의 기술 수용 정도와

관련해서, 몇몇 국가들은 이미 초기 수용자 (Early Adopter) 단계에서 초기 다수자 (Early Majority) 단계로 넘어가고 있다고 밝혔다.^① 이러한 리포트를 통해, 혁신 확산 이론 (Diffusion of Innovations) [17]을 기반으로, VUI 시장의 성숙도를 확인할 수 있으며, 동시에 다양한 사용자 층이 활발히 VUI를 수용하기 시작했음을 알 수 있다.

초기 다수자 (Early Majority)가 VUI 기술을 수용하게 되었다는 것은, 이제 VUI가 더 이상, 기술에 대한 호기심과 이해가 깊은 초기 수용자 (Early Adopter)의 전유물이 아니라, 더 넓은 사용자 층도 접할 수 있는 기술이자 인터페이스가 되었으므로 해석할 수 있다. 이는 VUI와 관련한 기술을 연구 및 개발하고 기기와 서비스를 만드는 주체에게도, 타겟으로 보아야 할 사용자 대상이 넓어지고 있음을 의미하기도 한다. 또한, 앞으로 시장에 출시하게 될, VUI를 적용한 새로운 기기 및 서비스는 여러 층위의 사용자를 출시 초기부터 맞아들이게 될 수 있다는 것을 의미한다.

한편, VUI를 구성하는 세부 기술이 빠르게 발전되고 있음에도 불구하고, 해당 인터페이스는 아직까지 사용자에게 다양한 형태의 어려움을 유발하고 있다 [27]. 예를 들면, VUI의 사용자 입력과 시스템 출력을 직접적으로 확인하기 어렵기 때문에 발생하는 문제가 많다 [26]. 음성만을 상호작용의 매개로 활용하는 Voice-Only 인터페이스에서 특히 많이 발생할 수 있는 문제인데, 사용자가 자신이 말한 것을 시스템이 정확히 이해했는지 확인하기 어려운 상황에 봉착하는 경우가 이에 해당한다.

VUI를 사용할 때 발생하는 다양한 어려움에 대해서, HCI 커뮤니티는 VUI 발전 초기부터 다양한 연구를 진행해 왔다 [40, 41]. 최근에도 VUI 이용 시 경험하는 어려움에 대한 연구는 지속해서 진행되고 있다 [3, 4, 6, 27, 29]. 그러나 다양한 사용자 층이 경험하는 어려움을 비교하는 연구는 많지 않다. 현재와 같은 VUI 확산의 시대에는 VUI를 이용하기 위한 진입

^① <https://voicebot.ai/2018/10/19/phase-one-of-the-voice-assistant-era-is-over-long-live-phase-two/>

장벽이 낮다. 그렇기 때문에, 기술 수용 관점에서 수용 적극성이 낮은 사용자도 VUI를 이용해볼 수 있다. 이는 같은 VUI를 쓰는 여러 사용자 그룹이 존재함을 의미하며, 이들이 경험하게 될 어려움에는 차이가 있을 것이라 생각해 볼 수 있다. 그러나 현재의 연구는 사용자 층에 대한 비교 연구를 통해 VUI 이용 시 맞닥뜨릴 어려움에 대한 다각도의 이해를 고려하지는 않는 듯 하다.

본 연구는, 이와 같은 시장 상황과 기술 수용도를 염두에 둔 채 사용자 층을 분류하고, 서로 다른 사용자 층이 VUI를 이용하면서 겪는 어려움에 대해 이해하는 것을 목표로 한다. 이를 통해, VUI 기술 연구 및 개발과 디자인, 나아가 서비스 출시에 있어, 각각의 사용자의 어려움에 대해 이해하고 이를 개선하기 위해 어떤 점을 고려해야 할지에 대해 살펴보고자 한다.

1-2. 연구의 내용 (Research Goal)

본 연구는 네이버 클로바 AI (Naver Clova AI)에서 자체 기술력으로 개발하고 디자인한, 식당 예약 음성 에이전트 AiCall을 활용하여^②, 사용자의 VUI 상호작용에 대해 이해하려 한다. 이 때, 사용자 그룹을 이용 빈도와 최신성을 기준으로 나누고 상이한 사용자 그룹이 어떻게 상호작용하는지에 대해 탐구하고자 한다.

1-2-1. 사용자 그룹 구분

혁신의 확산 (Diffusion of Innovation)에^③ 따른 기술 수용 관점에서

^② <https://clova.ai/aicontactcenter>

^③ 혁신의 확산은 새로운 아이디어와 기술이 어떻게, 왜, 어떤 속도로 퍼지는지를 설명하는 이론이며, Everett Rogers에 의해 발표되었다. 확산은 사회 시스템 내의 구성원 사이에서 시간이 경과함에 따라 혁신이 전달되는 과정이라고 주장한다. 혁신은 이를 받아들이는 과정에서 특정 지점 별, Critical Mass에 도달한다고 말하고 있다. 이 지점의 구분에 따라, 혁신을 받아들이는 과정이 혁신가 (Innovator), 초기 수용자 (Early Adopters), 초기 다수자 (Early Majority), 후

서로 다른 사용자 그룹에 대해 이해하기 위해서는, 먼저 사용자 그룹을 구분할 필요가 있다. VUI가 초기 다수자 (Early Majority)에게까지 받아들여지는 현황을 감안할 때, 본 연구에서는 다음의 가정을 적용하여 사용자 그룹을 구분하려 한다.

- 초기 다수자가 기술 (VUI)을 받아들일 수준으로 기술이 확산되었다면, 그것보다 더 기술에 대해 열려 있는 태도를 가졌을 혁신가 (Innovator), 그리고 초기 수용자 (Early Adopter) 그룹은 해당 기술을 활발히 활용하고 있을 것이다.
- 하지만, 혁신가 혹은 초기 수용자 역시 해당 기술을 일찍 접했다 할지라도, 수용의 과정에서 지속적으로 사용 하고 있을 것이라 단언할 수 없기 때문에 최근까지 해당 기술을 사용하고 있는지 확인할 필요가 있을 것이다.
- 이에 따라, 기술에 대한 수용도에 대해 사용자 층을 나누는 기준으로 이용 빈도 (Frequency of Use)와 이용 최신성 (Recency of Use)을 활용한다. 얼마나 활발히 기술을 이용하는지를 알아보는 것을 통해, 기술 수용도라는 원래의 취지에 맞는 사용자 구분을 가능케 할 것이다.
- 즉, 이용 빈도가 높고 동시에 이용 최신성이 높은 사용자 층은 VUI 수용 관점에서의 기술 수용도가 높은 사용자 층으로 가정한다. 이들은 초기 수용 시기에 인입되었던 사용자일 가능성이 높을 수 있다. 본 연구에서는 이를 Heavy User라 지칭한다.
- 이용 빈도가 낮거나 없고, 이용 최신성 역시 낮거나 없는 사용자 층은 기술 수용도가 낮은 사용자 층으로 가정한다. VUI 수용에 있어 수용 적극성이 상대적으로 낮은 초기 다수자를 대표하는 사용자 층으로 가정해 볼 수 있다. 본 연구에서는 이를 Light User라 지칭한다.

기 다수자 (Late Majority) 그리고 후발 주자 (Laggards)에게로 전파된다고 말한다.

이렇게 이용 빈도와 이용 최신성을 기준으로 사용자 그룹을 크게 두 개로 나누어, 서로 다른 사용자 층이 VUI를 어떻게 사용하는지에 대해서 알아보고자 한다.

1-2-2. 연구 질문 (Research Question)

본 연구는 VUI에 관련된 현재의 HCI 커뮤니티의 연구의 흐름 중, VUI를 활용하며 사용자들이 겪는 어려움과 관련된 연구와 그 궤를 함께하려 한다. VUI와 관련한 기술 그리고 디자인 방법론과 그것을 표현하는 기기 및 서비스까지 이 모든 것이 현재 완전한 성숙단계에 올라와 있다고 볼 수 없다. 그러므로, 필연적으로 사용자가 이를 사용하며 겪는 어려움 (Difficulties)은 존재할 수 밖에 없다. 인터페이스의 특성 상 그래픽 기반 사용자 인터페이스 (Graphical User Interface, GUI)의 성숙도에 비해 아직 낮은 VUI의 성숙도 역시 사용자로 하여금 GUI에 비해 더 많은, 또한 다른 형태의 어려움을 겪게 하는 주된 요인으로 작용한다. 이러한 환경 하에서, VUI 활용의 어려움에 초점을 맞추는 것은 현재의 기술 발전 단계에서 의미 있다.

본 연구에서는 이용 빈도와 최신성을 기준으로 구분한 두 사용자 그룹, Heavy User 그룹과 Light User 그룹을 대상으로 VUI 상호작용 환경에서 다음의 내용을 알아보고 탐구해보고자 한다.

1. Heavy User 그룹과 Light User 그룹은 새롭게 이용하게 되는 VUI에 대해 어떤 어려움을 겪을 것인가?
2. 상이한 사용자 그룹이 겪는 어려움은 어떠한 기술, 혹은 VUI 디자인에서 기인하는가?
3. 각각의 사용자 그룹이 VUI를 이용하며 겪게 되는 어려움에 대처하는 방식은 무엇인가?
4. 위 세 가지 질문을 통해 파악한 내용을 근거로, 상이한 사용자

그룹이 경험하는 어려움을 해결하기 위해 VUI는 어떤 방향으로 개선될 수 있는가?

본 연구는, 위 네 가지 주요 연구 질문을 해결하기 위해, 앞서 언급한 식당 예약 음성 에이전트 AiCall을 활용한 유저 스터디 (User Study)를 진행하고자 한다. 해당 유저 스터디에서 나온 정성적 (Qualitative), 정량적 (Quantitative) 결과를 분석하고, 사용자의 어려움을 개선할 수 있는 VUI 개발 혹은 디자인의 방향성에 대해 논의한다.

2. 관련 연구 (Related Work)

본 연구는, 크게 사용자 그룹 비교 연구이며 동시에 인터페이스를 사용하며 경험하는 어려움에 대한 연구이다. 본 장에서는 먼저, 본 연구에서 핵심적으로 다루는 상호작용 시 사용자가 겪는 어려움, 즉 대화의 관점에서 볼 때의 소통 실패(Communication Breakdown)와 관련된 기존 연구의 흐름에 대해 알아본다. 이를 통해, 본 연구가 해당 흐름에서 어떤 위치를 점하는지, 그리고 기존 연구와 어떻게 차별화되는지 밝히려고 한다. 두 번째로, Heavy User 와 Light User로 구분된 사용자 그룹에 대해 비교한 연구들을 조사한다. 마지막으로 VUI 적용 기기 및 서비스의 변화 흐름에 대해 알아보고, 그에 따른 연구 흐름에 대해 이해한다.

2-1. 소통 실패와 복귀 전략

소통 실패 (Communication Breakdown)는 VUI를 포함한 대화형 인터페이스에서 지속적으로 주요한 주제로 다뤄지고 있다. 일반적으로 대화에 있어 소통 실패는 대화 참여자 간 주고 받은 정보에 대한 서로 다른 이해가 발생함을 의미한다 [3]. 실제 사람 간의 대화에서는 이러한 소통 실패가 발생했을 때, 상호 간의 이해의 차이를 줄이기 위한 일련의 대화를 이어나간다. 이러한 상호 이해를 회복하기 위한 과정을 대화 수정 과정 (Repair Practice)이라 지칭한다 [35]. 대화에서의 상호 이해 회복 과정은 패턴화 시킬 수 있으나, 결코 단순하지 않다. 사람 간의 대화에 있어서의 소통 실패와 수정 과정을 포함한 복귀 전략은 대화 분석 (Conversation Analysis) 영역에서 패턴화하여 정리된 바 있다. 대화형 인터페이스를 구성하는 기술들이 발전되면서, HCI 영역에서도 소통 실패와 복귀 전략에 대한 다각도의 이해를 시도하는 움직임이 있다.

2-1-1. VUI 상에서의 소통 실패와 복귀 전략

먼저, 인터페이스 특성 관점에서 VUI를 포함한 대화형 인터페이스 (Conversational User Interface, CUI)와 그래픽 인터페이스 (Graphical User Interface, GUI)를 비교하는 연구들은 대화형 인터페이스가 가진 인터페이스 특수성에 집중한다 [14, 26]. Murad et al.은 [26] VUI는 사용자의 입력에 대한 처리 과정의 확인 가능성이 GUI보다 떨어지기 때문에, VUI 디자인 시 이에 중점을 두어야 한다고 말한다. Habler et al. [14] 역시 그들의 연구에서 Voice-Only VUI 환경이 구체적이고 섬세한 과제를 수행하는 환경에서 GUI에 비해 사용성이 낮다고 주장한다. 이들은 GUI와 VUI의 특성을 비교하며, 소통 실패를 유발하는 대화형 인터페이스의 본질적 제약에 대해 언급했다.

조금 더 구체적으로, 음성만을 소통의 매개로 활용하는 Voice-Only VUI의 경우 시스템이 현재 어떤 상태인지 알기가 어렵다. 예를 들어, 사용자가 의도한 발화가 시스템에 오류를 불러일으켰을 때, 사용자의 말을 들었는지, 들었는데 못 알아들었는지, 또는 알아들었는데, 답변이 잘못 매핑되었는지 정확하게 알기 어렵다. 결국 이는 사용자로 하여금 다음에 어떻게 행동해야 하는지 결정을 어렵게 하는 경향이 있다. 이처럼 VUI 상에서 상호작용은 소통의 실패가 생길 가능성이 높으며, 어떤 형태로 실패 지점에서 복귀 혹은 수정 (Repair) 해야할 지 알기 어려운 상황이 발생할 가능성이 높다.

VUI와의 상호작용 시 사용자가 소통 실패를 해결할 수 있는 방법은, 주로 대화를 통한 수정 과정 (Repair Practice)을 사용하는 것이다. HCI 커뮤니티에서는 지속적으로 이러한 대화형 인터페이스의 소통 실패 상황에 대한 수정 과정을 포함한 복귀 전략과 관련한 연구들을 진행해 왔다.

먼저 본 연구와 밀접한 관련이 있는 VUI에 관한 소통 실패와 수정 과정에 대한 연구들이 있다. Beneteau et al.은 [3] 그들의 연구에서 가족 단위로 Amazon Alexa를 사용할 때 소통에 문제가 발생하는 경우 이에 대해 사용자가 어떻게 대화를 수정하고 복귀하는지 파악했다. 해당 연구의

특징은 VUI의 주된 사용 맥락인 스마트 스피커, 특히 가정 내에서의 사용에 초점을 둔다. 그들은 1:1 상호작용이 아닌, 다수 사용자와 VUI 간 상호작용에 대해 초점을 맞추었고, 이러한 다수 사용자가 소통 실패에 대해 어떤 복귀 전략을 취하는지 연구했다. 또한, Myers et al.은 [27] 그들의 연구에서, 사용자가 Multi-modal 환경에서의 VUI와 어떻게 소통하고, 어떤 사용 상의 장벽을 맞이하는지에 대해 연구를 진행했다. 해당 연구는 사용자가 맞이하게 되는 장벽의 종류가 무엇이 있는지 알아보았고, 사용자들의 극복 전략에 대해 탐구했다. Cheng et al.은 [4] 미취학 아동이 VUI 상에서의 소통 실패에 대한 복귀 전략을 어떻게 펼치는지 연구했다. 그들은 어린이들이 복귀 전략을 사용하는 특징에 대해 비교적 다각도로 조사했다. 또한 Beneteau et al.의 연구와 유사하게 이 연구도 다수가 함께 복귀 전략을 사용할 때의 특징에 대해 연구했다. 구체적으로 어린이와 VUI가 1:1로 상호작용 할 때 발생하는 소통 실패에 대응하는 상황과, 부모와 함께 있을 때의 복귀 전략 방식의 차이 등도 조사했다.

위 연구에서처럼 소통 실패와 복귀 전략을 연구의 중심 주제로 선정하지 않은 경우에도, 소통 실패는 VUI 연구에서 기본적으로 다뤄진다. 먼저 Luger et al.은 [23] 그들의 연구에서, 사용자들이 에이전트에 대해 기대하는 바가 실제적인 소통 실패를 경험한 이후 어떻게 변해가는지 다루었다. 결국 사용자 기대감에 변화를 이끄는 주 요인은 소통 실패라는 주제와 깊이 연관되어 있음을 이 연구 결과를 통해 알 수 있다. Porcheron et al.은 [29] Amazon Alexa를 사용하는 가족들을 대상으로 한 연구에서 그들의 일상 속에서 Alexa가 어떤 역할을 하는지 다루었다. Alexa와의 상호작용에서 에이전트가 실패하는 경우, 가족들이 어떻게 복귀 전략을 전개하는지에 대한 양상을 통해, VUI 상에서의 에이전트 발화 또는 상호작용에 대한 제언을 한다. Clark et al.은 [6] 소통 실패와 복귀 전략을 다루는 것에서 더 나아가, 잦은 소통 실패는 현재 대화형 인터페이스와의 상호작용에서 필연적인 상황임을 가정한다. 이러한 가정 하에서, 사용자가 사람과의 대화처럼 복귀 전략을 펼치지 않을 수 있도록 시스템이 사용자를 안내해줄 필요가 있음을 제안한다. 이처럼 VUI를 대상으로 한 여러 유저

스터디에서 소통 실패는 중요한 주제 중 하나로 다뤄지고 있다.

2-1-2. 챗 인터페이스 상에서의 소통 실패와 복귀 전략

대화형 인터페이스 중, 챗봇 (Chatbot)으로 불리는 챗 인터페이스 (Chat Interface) 상의 소통 실패와 복귀 전략에 대한 연구도 지속적으로 진행되고 있다. 챗 인터페이스는 대화를 통해 상호작용이 일어나는 인터페이스라는 점에서 VUI와 공통점이 있다. 동시에 GUI의 인터페이스 상의 장점도 수용할 수 있다는 점에서 챗 인터페이스는 VUI와 차별점을 지닌다. 공통점과 차이점이 존재함에도 불구하고, 챗 인터페이스의 소통 실패와 복귀 전략에 대한 이해는 VUI의 그것에 대한 이해 및 적용에 도움이 될 수 있다.

챗 인터페이스에서 발생할 수 있는 소통 실패의 형태는 VUI에서의 그것과 유사하다. 하지만, 복귀 전략에 있어서는 GUI 상에서 선택할 수 있는 다른 부가적인 요소들, 이를테면 버튼 형태의 객관식 항목이나, 하이퍼 링크 등을 활용한다는 측면에서 VUI와 차이점이 있다. Ashktorab et al.은 [1] 그들의 연구에서 챗 인터페이스에서 소통 실패에 대한 다양한 복귀 전략에 대해 비교 연구했다. 대화형 인터페이스의 장점 뿐 아니라, GUI의 장점을 활용한 방향으로 소통 실패를 표현하고, 복귀 전략을 다양화할 수 있다는 점은 챗 인터페이스의 특성 덕분에 가능하다. Folstad et al.은 [12] 그들의 연구에서 실제 운용되고 있는 챗봇 서비스에서 발생하는 소통 실패와 관련된 문제들과 복귀 전략에 대해 다루었다. 소통 실패를 해결하기 위해 화면 상에서 사용자가 선택할 수 있는 ‘대화 제안’ 기능을 적용했을 때, 어떤 효과가 실제로 있는지 알아보았다. 그 결과 ‘대화 제안’ 기능이 소통 실패가 확장되는 것을 방지함을 확인했다. 챗 인터페이스의 복귀 전략에 관련한 연구는 GUI의 특징을 반영하는 경우가 많다. VUI가 가진 음성 중심의 상호작용에서는 위와 같은 복귀에 대한 전략적 방안을 적용하기 어려운 부분이 많으며, 부분적으로만 참고할 수 있다.

챗 인터페이스에서 발생하는 소통 실패의 기술적 원인은 VUI의 그것보다 비교적 단순하다. VUI의 경우, 사용자의 음성 입력을 1차적으로 처리하는 음성 인식 (ASR)부터 에러가 발생할 수 있다. 또한 VUI에서는 음성 합성 (TTS)을 통해 전달되는 에이전트의 목소리 또한 소통 실패의 원인이 될 수 있다. 하지만, 챗 인터페이스의 경우 비교적 정확하게 사용자 입력을 받아들일 수 있고, 시각으로 확인 가능한 자연어 발화를 통해 에이전트의 의도를 전달하는 경우라면, 시스템 출력과 관련한 오류 가능성 역시 적다. 이러한 이유로, 챗 인터페이스의 오류 상황에 대한 기술적인 이해와 발전 방안에 대한 논의 역시 자연어 이해 (NLU)와 대화 운영 (DM) 영역에 집중되는 경향이 크다 [34, 42]. 하지만 VUI에서 발생하는 소통 실패와 관련된 기술적 요소는 이것보다 다양하며, 사용자 측면에서 실패라고 느끼게 되는 요소 역시 복합적이다. 이러한 의미에서 VUI에 대한 소통 실패와 그에 대한 사용자의 복귀 전략에 대해 알아보는 것은 기술적으로 보았을 때 VUI가 사람과의 소통과 유사한 내추럴 인터페이스 (Natural Interface)로 발전하는 데에 도움을 줄 것이다. 이는 사용자 관점에서, 직관적 인터페이스로서의 VUI의 효용을 높여줄 수 있다.

본 연구는, VUI에서 발생하는 소통 실패에 대해 다룬다는 점에서 위의 연구와 맥락을 같이한다. 하지만 차이점은 사용자 그룹을 분리한 비교 연구라는 점에 있다. 상이한 사용자 그룹이 VUI를 사용하면서 맞닥뜨리게 될 어려움에 대해 이해하는 것을 넘어, 각 사용자 그룹이 보고한 어려움에는 어떤 차이가 있고, 무엇을 유의미한 차이로 볼 수 있는지 밝혀낸다. 나아가, 이들이 어려움을 극복해나가는 과정에서 어떤 형태의 반응, 태도 혹은 전략을 취하는지 알아보고, 이러한 사용자의 특징을 감안할 때 VUI가 고려해야 할 것은 무엇인지 알아본다. VUI에 관한 주요 연구 주제인 소통 실패에 대한 큰 맥락을 함께 가져가되, 사용자 그룹을 달리하여 연구를 진행한다는 측면에서 본 연구는 이전의 연구들과 차별성을 가진다.

2-2. Heavy User와 Light User 사용자 그룹 비교 연구

Heavy User와 Light User의 비교는 사용자 행동 패턴 분석을 통해 논의 주제나 결과를 도출하는 연구에 주로 활용된다 [10, 11, 28]. 이러한 비교 연구의 목적은 일반적으로 사용자 (또는 소비자)를 깊이 이해하고, 그들에게 적합한 제안, 처치나 기여를 하기 위함이다. Heavy User와 Light User의 비교를 활발히 이용하는 분야 중에는 중독적 성향에 대한 심리적 치료에 대한 분야 [11, 37], 그리고 소비 패턴 분석에 활용하기 위한 마케팅 및 경영학 분야 [13, 36]등이 있다.

HCI 연구에서도 Heavy User와 Light User를 나누어 사용자를 이해하는 형태의 연구들이 있다 [31, 38]. Shim et al. 은 [31] 자신들의 연구에서 기술 수용 모형 (Technology Acceptance Model, TAM)에 [20] 근거하여 모바일 앱을 사용하는 사용자 그룹 별 동인이 어떻게 다른지 분석한 바 있다. Verkasalo et al. 역시 [38] 그들의 연구에서 기술 수용 모형을 사용하여 사용자들이 해당 스마트폰 어플리케이션을 사용하게 만드는 동인과 방해하는 요인에 대해 연구한 바 있다. Drabble 은 [10] 자신의 연구에서 사회 관계망 서비스 (Social Networking Service, SNS)의 Heavy User와 Light User를 나누고 이들의 부가적인 SNS 사용이 실제 교육에 어느 정도 영향을 끼치는지 연구한 바 있다.

이렇게 보았을 때, Heavy User와 Light User를 나누어 비교하는 연구의 흐름은 크게 둘로 나누어 볼 수 있다. 그 하나는, 그들이 왜 Heavy User 혹은 Light User가 되었는지 이해하는 것이다. 기술 수용 모델을 활용한 각 사용자 그룹에 대한 이해는, 기술에 대한 그들의 긍정적이거나 부정적인 동인을 이해하는 데에 초점을 맞춘다. 특히 중독 치료 연구에서 역시 그들이 왜 Heavy User가 되었는지 혹은 Light User가 되었는지에 집중한다 [39]. 또 하나의 흐름은 분류된 사용자 그룹의 표현된 행동 자체를 이해하고 파악하는 것이다 [18, 33].

본 연구는 위에서 기 분류된 상이한 사용자 그룹의 행동 패턴을 이해하는 것을 그 목적으로 한다. VUI 상에서의 Heavy User와 Light User를 나누는 기존 저서의 분류 기준을 근거로 [7] 이들이 VUI를 이용하는 행동 패턴의 차이에 대한 이해를 하는 것이 본 연구의 주된 목적이다. 이러한 사용자에게 대한 이해는 VUI 기술 관점에서의 개선 사항을 도출에 필요한 논의를 이끌어나가는 데 주요한 역할을 할 수 있을 것이다.

2-3. VUI 적용 기술 및 제품 그리고 서비스의 발전 흐름

음성 인식 (ASR), 음성 합성 (TTS) 등의 음성 관련 기반 기술은 20세기 초반부터 연구되기 시작했지만, 20세기 중반이 지나서야 그 실용성을 갖추기 시작하며 세상에 소개되었다. 1952년, Bell Labs의 Audrey라는 Speech Recognizer를 시작으로, 음성 관련 기반 기술은 꾸준히 발전을 거듭해왔다. 기술의 성능과 비용의 효율화가 지속해서 이뤄지면서, 20세기 후반 음성 관련 기반 기술은 일반인들에게 제공 가능한 VUI의 형태로 세상에 본격적으로 선보이게 되었다 [24].

2-3-1. 대화식 음성 응답 (Interactive Voice Response, IVR)

VUI가 세계적으로 활발하게 사용되기 시작한 시점은, 2000년대부터라고 할 수 있다. 국내에서는, 본격적으로 적용되었다고 볼 수 없지만, 해외에서는 대화식 음성 응답 (Interactive Voice Response, IVR)이 본격적으로 산업계에서 활용되기 시작하였다. IVR은 음성 인식 기술과 이중 톤 다중 주파수 (Dual Tone Multiple Frequency, DTMF)요소를 활용한 기술이자 인터페이스이며, 전화 통화 맥락에서 활용되었다.^④ 주로 전문적인 전화 응대나 상담이 필요한 영역에 IVR을 활용한 서비스가 많이 적용되었으며, 주로 금융, 여행, 문화 산업 등에서 활용되었다.

IVR 기술은 전화 응대가 필요한 산업 영역에서 제한적이지만

^④ https://en.wikipedia.org/wiki/Interactive_voice_response

지속적으로 활용되고 있으나, 사람들은 IVR의 음성을 통한 상호작용 (Voice Interaction)은 불편한 것으로 인지하고 있다. 그 이유는 과거의 불편한 사용 경험이 축적되었기 때문이다. IVR을 도입한 산업계는 서비스 상 정확한 목적성이 있는, 목적 지향 대화를 기본으로 한다. 즉, 단순하고 제한된 형태의 대화 흐름만이 용인되는 것이다. 이에 더해 과거 음성 관련 기술의 제한적 발전 상황은 정의된 전화 흐름 (Call Flow)을 벗어나는 사용자의 발화를 인식하고 처리하지 못하는 제약이 있었다.

위와 같은 서비스 목적 상, 그리고 기술 상의 제약을 벗어나기 위해, 전화 맥락에서의 VUI 디자인에 대한 다양한 연구가 진행되었고 관련 저서 등이 출판되었다. Clark et al.의 조사 연구에 따르면 [5], IVR에 대한 사용 경험을 중심으로 연구한 몇몇 연구가 있고, 주로 앞서 언급한 IVR의 활발한 적용 시기였던 2000년대에 진행되었다. Wolters et al.은 [41] 인지 부하 중 Working Memory Load 관점에서 연구를 진행했다. 전화 통화 맥락이 사용자의 음성 상호작용에만 의존하는 특성을 이해하고, 전화의 목적을 달성하는 과정에서 인지 부하가 어떤 영향을 주는지 파악했다. Wilkie et al.은 [40] 전화 통화 중 사용자에게 정보를 제공하는 방식을 탐구하는 측면에서 연구를 진행했다. 어떤 형태의 정보 제공 방식이 사용자에게 더 효용이 있는지를 살펴본 연구였다. 이와 같은 연구들은 실제로 제한된 형태의 상호작용만이 가능한 IVR 환경 하의 VUI에서 사용 경험을 높이하고자 하는 측면에서 진행되었다고 볼 수 있다.

또한 IVR이 활발하게 적용되었던 시기에, VUI에 대해 전반적인 가이드라인 및 매뉴얼을 제공하고자 했던 여러 저서 등이 출판되었다. Cohen et al. 은 [7] IVR 환경의 VUI를 디자인하기 위해 고려할 여러 사항에 대해서 정리하였다. 주지할 점은, 이 때의 VUI 디자인 가이드라인은 현재의 VUI 디자인에서도 충분히 유효하게 적용할 수 있다는 점이다. 물론, 음성 인식 성능 및 NLU 성능의 향상을 위해 Grammar를 정교하게 구성한다거나, 음성 합성의 제약을 보완하기 위해 녹음을 하는 방식 등을 소개하는 것을 보았을 때, 현재의 기술에서는 고려하지 않아도 될 내용이

포함되어 있는 것은 맞다. 하지만 그림에도 불구하고, 아직도 디자인 측면에서 고려해야 할 사항이 요점 별로 정리되어 있다. 또한, Harris 역시 VUI를 디자인하기 위해 고려해야 할 여러 사항과 디자인 프로세스를 정리해 놓은 저서를 출판하였고 [16], 현재 새로운 VUI 시대가 진행되는 상황에서도 그 디자인 프로세스나 업무 방향성은 여전히 유효하다.

과거 VUI가 전화 통화 맥락인 IVR 기술 하에서 제한적으로 사용되던 시대에도, 사용자와의 상호작용을 보완하기 위해 고려할 사항들은 많았다. 오히려 기술적 제약이 현재보다 분명히 존재하던 시기였기 때문에, 상호작용에 대한 디자인에 더욱 신경을 써야했을 수 있다. 음성 관련 기술이 비약적으로 발달되고, 관련하여 적용할 수 있는 서비스 또는 기기의 범위 역시 아주 많이 늘어난 최근에도 과거의 연구 방향성과 가이드 혹은 매뉴얼은 유효하다.

2-3-2. 스마트 디바이스의 VUI

최근, VUI는 스마트 디바이스의 등장과 함께 다시 한 번 확산되는 시기를 맞고 있다. 2011년 Apple의 Siri가 iPhone에 탑재되면서, 지능형 가상 비서 (Intelligent Personal Assistant, IPA)의 시대가 본격적으로 도래했다. 사용자는 Siri에게 음성 명령을 내림으로써 스마트폰 내의 간단한 태스크를 수행할 수 있고, 음성 피드백 혹은 화면을 통한 피드백을 받을 수 있다. IVR의 확산에 따른 2000년대의 양상과 차이가 나는 점은, 스마트폰이라는 개인 기기에 탑재된 VUI가 각 개인의 가상 비서 역할을 했다는 점이다. 실제로, IPA는 과거의 목적 지향적인 대화 뿐 아니라, 다양한 형태의 음성 상호작용을 커버해야 하고 나아가 일상 대화까지 일부 소화하는 면모를 보여야 했다. 이와 같은 VUI의 확산은 IPA에 초점을 맞춘 HCI 커뮤니티의 연구가 진행되는 환경적 토대를 만들었다. Luger et al.은 스마트폰에 탑재된 VUI, 대화형 에이전트에 대한 사용자의 기대감과 현실 사이의 괴리에 대해 연구를 진행했다 [23]. 해당 연구는, 사용자들이 스마트폰에서 활용 가능한 대화형 에이전트에 대해 탐색적 일상 대화를

통해 단순한 인터페이스 이상의 추상적 기대감과 의인화를 진행하는 경향이 있음을 보여주었다.

또한, 2014년 11월 Amazon의 Echo라는 스마트 스피커가 출시되며, VUI는 개인용 스마트 디바이스에서, 활용 공간의 개념적 확장을 이뤄냈다. 주로 개인 주거 공간에 비치된 스마트 스피커의 사용 맥락이 확대되었다. 스마트폰의 VUI가 주로 개인과의 1:1 대화 혹은 기능 수행에 초점이 맞추어져 있다면, 스마트 스피커의 VUI는 다수의 사용자와 대화하거나 복수의 사용자 이용 맥락에서 기능을 수행하는 것도 고려되어야 했다. 스마트 스피커가 등장한 초기, 이와 관련한 연구는, 스마트폰의 VUI, 즉 지능형 개인 비서와의 상호작용과 의인화에 대한 맥락을 이어가는 경우가 있었다. Purington et al.은 [30] Amazon의 스마트 스피커 Echo와 가상 비서 Alexa의 사용 리뷰를 활용해서, 사용자들이 해당 기기에 대해서, 구체적으로 Alexa라는 가상 비서에 대해서 어떻게 인지하고 있는지에 대해 분석하였다. Lopatovska & Williams도 [22] 이와 유사하게 유저 스터디를 수행하고, 참여자로부터 Amazon Alexa에 대해서 어떤 감정을 느끼게 되는지, 그리고 그 이유가 되는 상호작용 배경은 무엇인지 연구하였다. VUI가 목소리로 소통하는 인터페이스, 나아가 목소리를 가진 인터페이스라는 특징은 사용자로 하여금 VUI에 대한 의인화를 시킬 가능성을 가지고 있었다. 이런 인터페이스의 특징에, 발전된 인공 지능 기술과 스마트 디바이스 상에서의 다양한 적용 영역이 만나게 되면서, VUI는 의인화된 지능적 개체로 인지되기 시작했다.

2-3-3. AiCall의 기술 및 적용 영역 상의 위치

AiCall은 전화 상의 VUI로, IVR 기술 영역과 많은 부분 겹쳐 있다. 그러나, 과거 IVR이 기술 상의 이유로, 단순하고 제한적인 전화 흐름만을 다루었던 반면, AiCall은 발전된 인공 지능 기술로 과거의 제한적 요소를 해소하고자 했다. AiCall을 활용한 본 연구를 통해 전화 상의 VUI와 사용자가 어떻게 상호작용하는지 알아볼 수 있다. 발전된 기술로 재정비된

전화 상의 VUI라 할지라도, 전화 환경이라는 특수성이 사용자에게 미치는 영향이 있는지 역시 확인할 수 있다. 특히 본 연구의 주제인 사용자의 어려움, 즉 소통 실패에 전화 환경이라는 특수성이 영향을 미치는지에 대해 확인할 수 있다.

3. 연구 방법 (Methodology)

본 연구는, 식당 예약 전화 에이전트 AiCall을 활용해, 서로 다른 사용자 그룹이 VUI를 활용할 때 겪는 어려움을 파악한다. 그리고 그들이 어떻게 해당 어려움에 반응하고 대응하는지 확인한다. 이를 통해, VUI 기술 요소의 개선 방향성 및 디자인 개선 방향성에 대해 논의한다. 이를 위해 식당 예약 전화 에이전트 AiCall이 어떻게 디자인, 개발 그리고 제작의 과정을 거쳤는지 소개하고, 사용자 조사를 진행한 방식에 대해 본 장에서 기술한다.

3-1. AiCall

AiCall은 식당 예약 전화 에이전트이다. 사용자는 AI 상담원과의 통화를 통해, 식사를 예약할 수 있고, 예약 정보를 조회 또는 변경할 수 있으며 취소할 수 있다. 또한, 전화를 통해 간단한 질문 (Frequently Asked Question, FAQ)에 대한 답변도 들을 수 있다. 현재 시점 기준으로 아웃백 스테이크하우스 미금점에 실제 적용되어 있으며, 아래 홈페이지^⑤ 에서도 가상 아웃백 스테이크하우스 미금점 형태로 체험해볼 수 있다.

3-1-1. AiCall 시스템의 구성

AiCall은 VUI를 전화 환경에 구축한 제품이다. 시스템 구성의 개념도는 다음과 같다 (그림 1 참조).

^⑤ <https://clova.ai/aicontactcenter>

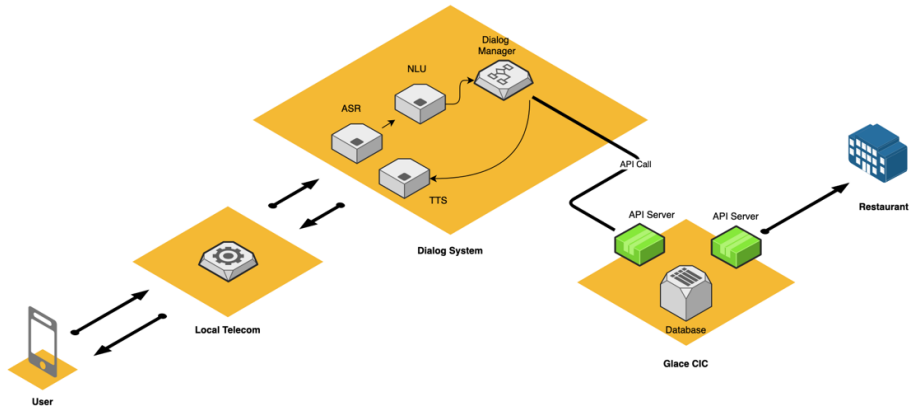


그림 1. AiCall의 구조 (업무 상의 비밀 유지를 위해 개념도로 표현)

먼저, 사용자가 AiCall에 전화를 하면, 지역 전화 사업자와 Clova 인터페이스의 통신을 통해 사용자는 대화형 에이전트와 상호작용을 시작할 수 있다. 사용자의 음성은 Clova의 음성 인식을 통해 텍스트로 이루어진 결과로 나오며, 해당 결과는 자연어 이해를 위한 입력값이 된다. 자연어 이해 모델은 이 입력값을 사용자 intent 형태로 출력하며, 해당 값은 Dialog Manager에 입력값으로 작용하며, 기 정의된 운영 방식을 통해 대화가 운영된다. 이렇게 결정된 대화의 결과값은 텍스트 형태로 음성 합성을 통해 출력된다. 즉, 대화형 에이전트가 정의된 발화를 출력하는 방식이다.

Dialog Manager는 기 정의된 대화의 정책적인 요소 역시 관리하는데, 그 중 서비스 주체인 Glace CIC, 즉 식당 예약 도메인을 관리하는 사업자와 API 단위의 통신을 통해 정보를 수렴하고 정책에 따른 결과를 출력한다. 이때, 식당 예약 도메인 관리자의 정보값에 따라, 대화형 에이전트의 서비스 커버리지가 달라질 수 있다.

대화형 시스템, 나아가 VUI가 서비스 영역에 적용되어 시스템이

구성되는 데에는 다양한 기술, 혹은 서비스 단위들이 결합되어야 한다. 그 중, 본 연구에서는 대화형 시스템 본연의 디자인과 개발에 대해서 조금 더 상세히 소개하고자 한다.

3-1-2. AiCall 대화 디자인

AiCall의 대화 디자인은 대화 분석 (Conversation Analysis, CA)이라는 학문 분야를 기반으로 삼아 설계되었다. 실제로 인간 - 인간 대화의 방향성이 HCI 영역에서 다루는 인간 - 기계의 대화와는 다를 수 있다. 대화 분석에서 정리된 결과를 활용하지만, 그것을 재해석하는 과정이 반드시 필요하다. 또한, 재해석하는 과정에서 사용자로 하여금, 자연스러운 대화가 일어난다는 사용 경험을 주기 위한 방법에도 초점을 맞추어 디자인, 크게는 설계 작업에 대한 진행이 필요하다. 본 연구에서는 그 개략적인 방향성을 언급하고자 한다.

3-1-2-1. TSAT (Turn - Sequence - Activity - Task)

AiCall 대화 설계의 근간은 TSAT 프레임워크에 기반하고 있다. TSAT 프레임워크는 Conversation Analysis 학문 분야의 대화에 대한 기초 단위 구분에 근거하여 [25] Clova AI DUET 팀 내부에서 자체 설계되었다. TSAT는 그 순서대로 Turn, Sequence, Activity 그리고 Task를 말한다. 디자인의 단위를 가장 작은 것부터 가장 큰 것까지 범주화시킨 것이라 볼 수 있다. VUI, 크게는 대화형 유저 인터페이스 (Conversational User Interface, CUI)의 대화 설계는 Turn단위의 설계에만 신경을 쓰는 경우가 많다. 즉, User turn 그리고 System turn의 나열로 대화를 만들어내는 방향성을 가지고 있다는 것이다. 하지만, 대화는 대화 참여자 간 발생하는 쌍방향 의사 소통이 겹으로 드러나는 현상이며 이 의사 소통 과정에서 대화 디자인, 혹은 더 큰 개념으로, 설계자가 예측한 하나의 대화 흐름 (Flow)이 아닌, 다양한 방식의 고려하지 못한 대화가 일어날 수 있다.

TSAT 프레임워크는 이러한 단선적인 대화 흐름 설계에서 기본적으로 벗어나 있다. Turn의 연결은 곧, Sequence라는 더 큰 단위로 결합된다. Sequence는 하나의 System turn과 하나의 User turn을 결합한 단위이며, 이 단위는 대화 설계의 중심축이 된다. 즉 Sequence 단위들의 결합으로 최소 단위의 대화라고 볼 수 있는 결과물이 만들어질 수 있는 것이다. AiCall 아웃백 스테이크하우스점 제품의 대화를 Sequence 단위로 나눠보면 다음과 같다.

S1: 언제로 예약 도와드릴까요?

U1: 내일 저녁 8시에 가려고 하는데요.

S2: 혹시 몇 분이서 오시나요?

U2: 4명이서 가려구요.

위 대화에서 S는 System 발화를 의미하고, U는 User 발화를 의미한다. S1 - U1을 하나의 Sequence로 이해할 수 있으며, S2 - U2를 또 하나의 Sequence로 이해할 수 있다. 이렇게 Sequence를 기준으로 대화가 진행되는 것을 확인할 수 있다

Activity의 개념은 대화 분석의 Conversational Activity의 의미를 차용하여 대화 설계에 적용하였다. 기본적으로 Activity의 개념은 인간 - 기계 대화에서 명명하는 하나의 카테고리인 목적 지향형 대화에서 유효하게 활용될 수 있다. 뚜렷한 목적이 있는 서비스 상황에서 활용될 수 있는 개념으로, 대화형 에이전트가 질문의 주도권을 가지며, 사용자가 목표한 바를 이뤄나갈 수 있도록 이끌어어나가는 형태를 취한다. 이런 경우, 에이전트가 사용자에게 질문을 진행하고, 필요한 정보를 사용자로부터 얻어내 대화 참여자 (에이전트 - 사용자) 모두가 가진 목적을 이뤄내는 것을 목표로 한다. 기본적으로 대화의 흐름이 존재하게 되는데, 이 때 유용하게 쓸 수 있는 개념이 Activity이다.

Activity는 Sequence 단위를 포함하는 상위 개념으로, 하나의 완결된,

최소한의 목표한 바를 담을 수 있는 개념적 단위이다. 즉, 대화를 통해 하나의 완결된 활동을 할 때 고려해야 할 대화 설계 요소를 포함한 것을 Activity 단위라고 할 수 있으며, AiCall 아웃백 스테이크하우스점은 개념적으로 다음과 같은 Activity를 가진다.

- Opening
- Scheduling
- Availability Check
- Confirmation
- Closing

Opening은 전화 대화의 시작점으로 인사를 진행하고, 사용자의 발화를 유도한다. Scheduling은 식당을 대상으로 전화 예약을 진행하기 위한 기본 정보를 대화형 에이전트가 받아나가는 과정이 디자인되어 있다. 즉, 날짜, 인원 및 시간 등 예약에 필요한 정보를 받는 Activity다. Availability Check는 실제로 고객이 제공한 정보를 매장에 전달한 경우, 매장이 이것을 처리할 수 있는지에 대한 확인과 관련된 과정, 그리고 다른 일정에 대한 제안 과정이 포함된 Activity다. Confirmation은 사용자가 원하는 일자에 예약이 가능한 경우, 그 확정 과정을 포함하며, Closing은 상호 간의 합의 과정을 통해 전화를 종료하는 흐름이 설계가 되어 있는 Activity다. 이 밖에도, AiCall은 위 흐름을 가능하게 하기 위한 여러 추가적인 Activity 단위가 동시에 존재하고 있다.

마지막으로, Task는 TSAT 프레임워크의 가장 상위의 개념적 단위이다. 사용자가 전화를 거는 상위 목적이 Task가 다루는 범위이다. 위의 다섯 가지 Activity는 신규 예약이라는 상위 Task의 세부 Activity들이다. AiCall은 기본적으로 이 외에도 다양한 Task를 가지고 있다. Task의 종류는 다음과 같다.

- RM (Reservation Make)

- RCG (Reservation Change)
- RCK (Reservation Check)
- RCL (Reservation Cancel)
- FAQ (Frequently Asked Question)

이와 같이 Task가 개념적으로 구성되어 있으며, 각 Task 내에는 복수의 Activity들이 포함되어 있다. 이로써 TSAT 프레임워크는 AiCall 전체 대화를 구성하는 핵심 요소로 적용되어 있다. (그림 2 확인)

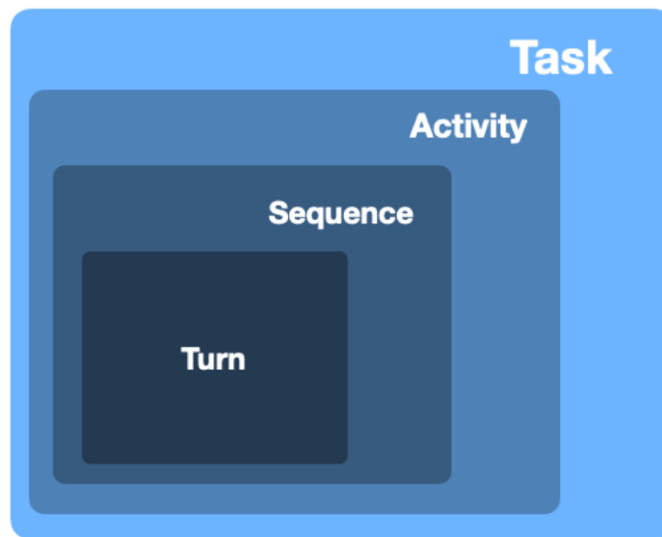


그림 2. TSAT 프레임워크의 관계적 개념도

TSAT 프레임워크는 위와 같이 복잡한 형태의 대화를 디자인하기에 용이하도록 파편화된 대화 단위를 그룹핑하는 데에 유용하다. 하지만 위 프레임워크로 작업을 진행할 때 그룹핑되어 있는 단위 사이를 대화가 이동하게 되는 경우 발생할 수 있는 Verbosity는 결과적인 문제를 야기할 가능성은 존재한다.

3-1-3. AiCall 에이전트 페르소나 디자인

AiCall의 페르소나 (System Persona)는 캐주얼 레스토랑, 혹은 패밀리

레스토랑의 점원을 대신할 수 있는 가상 에이전트를 염두에 두고 디자인되었다. 페르소나 디자인의 목표는 매장에서 일상적으로 들을 수 있는 언어를 사용하며 친절한 목소리로 응대하는 에이전트를 만드는 것이었다. 이를 위해서는, 해당 레스토랑의 다양한 지점의 고객 응대 통화 데이터를 통해, 응대의 우수 사례를 조사하고, 이를 적용할 필요성이 있었다. 이와 같은 목적으로 해당 작업을 특정 기간 동안 진행했다. (2019년 5월 초 ~ 6월 초, 1달 간)

VUI 상의 에이전트 페르소나 디자인의 핵심 요소인 목소리와 언어 스타일은 음성 합성 화자의 특질과도 깊은 연관이 있다. IVR을 중심으로 한 VUI의 1차적 활황기에는 오히려 제품 별, 혹은 서비스 별 특징에 따라 성우 녹음 스타일을 가이드하는 업무가 중요했다. 물론, 친절한 대화형 에이전트라는 이미지가 지배적으로 통용되었기에, 큰 틀을 벗어나지 못했지만, VUI 디자이너의 업무 영역에 녹음 스타일 가이드는 중요한 부분을 차지했다. 현재의 VUI 확산기는 스마트 기기에서 활용되는 대표 목소리가 중심이 되는 현상이 나타나고 있어, 오히려 섬세한 목소리의 디자인이 어려워지고 있다. 그럼에도 불구하고, AiCall 페르소나 디자인은 젊은 레스토랑 점원의 친절한 목소리와 매칭할 수 있는 여러 화자 후보를 고르고, 그 중 적합한 화자를 찾아 녹음 스타일을 가이드하는 과정을 거쳤다. 특수 목적의 음성 합성 화자이기 때문에, 많은 시간을 녹음에 할애할 수 없었으나, Clova Voice 팀의 음성 합성 기술을 활용해 적은 시간으로도 고품질의 합성음을 만들어낼 수 있었다 [32].^⑥

에이전트의 언어 활용 특징은 간접 높임법을 사용한다는 것이다. 서비스 직군에 있는 응대원들의 특이사항이라고 할 수 있는 높임법의 과한 사용은 사회적인 문제로까지 인식되고 있다.^⑦ 하지만, 역으로 기본적인 높임법에 대한 이해 부족으로, 간접 높임을 정확히 구사하는 경우에도

^⑥<https://clova.ai/voice?lang=ko>

^⑦<https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=103&oid=144&aid=0000277213>

높임법의 잘못된 사용이라고 생각하는 경향도 있다. 실제로, 간접 높임의 정확한 사용은 언어 사용에 있어 고객 응대의 질을 높여주는 역할을 한다. AiCall의 에이전트는 간접 높임의 활용을 통해, 사용자로 하여금 매장에서 서비스 응대하는 직원과 소통할 때의 느낌을 불러 일으키도록, 즉 사용자의 멘탈 모델 (Mental model)을 활용할 수 있는 방향으로 디자인되었다.

3-1-4. AiCall 데이터 구축 및 자연어 이해

AiCall은 자연어 이해 (Natural Language Understanding, NLU)의 성능 향상을 위해 자연어 데이터 구축에 노력을 기울였다. 182개의 User intent를 정의하고, 자연어 데이터를 모아 레이블링을 하는 과정을 거쳤다. 구축된 User Intent의 카테고리는 크게, 12가지이다. 12가지의 카테고리는 식당 예약, 특히 캐주얼 레스토랑 사용 상황에 적합한 형태를 추린 것이다. 각 카테고리 내에 정의된 서비스 요구 사항 뿐 아니라 대화의 진행 상황을 고려하여 구축된 User Intent 역시 존재한다.

자연어 이해는 BERT (Bidirectional Encoder Representation from Transformer, BERT) 모델을 Pre-trained하여 활용했다 [9]. 모델의 pre-train 방식으로 일반적인 대화 데이터 뿐 아니라, 유사한 질문 답변 영역의 다양한 데이터 역시 활용했다. 또한, 대화 맥락의 정확한 추론을 위해서, 학습을 위한 데이터의 구조를 다양화하고, 하이퍼 파라미터를 조정하는 과정을 거쳐 맥락을 이해한 답변이 가능하도록 했다. 또한, 음성 인식 결과를 활용한 데이터 학습을 통해 성능 향상에 기여하도록 했다.^⑧

3-1-5. AiCall 전화 환경 음성 인식

AiCall은 음성 인식 (Automatic Speech Recognition, ASR)에 있어, 전화망 음성 인식을 기본으로 하였다. 전화망 음성 인식은, 인터넷 망을 활용한 음성 인식에 비해, 낮은 샘플링 레이트 (Sampling Rate)를 활용한

^⑧ 자세한 내용은 보안 상의 이유로 밝히지 못함에 양해를 구함.

결과물을 음성 인식에 사용해야 한다. 8000Hz 샘플링 레이트는 인식 품질에 영향을 미칠 수 있으나, Clova Speech 팀의 기술력은 이러한 상황을 극복해내고, 높은 품질의 음성 인식 성능을 제공하였다.^⑨ 전화 환경은 이런 기본적인 샘플링 레이트 문제 뿐 아니라, 고려해야 할 사항이 많다. 특히 주변 소음에 강건한 음성 인식이 가능하기 위한 고민이 많이 필요하다. 통화를 하는 환경은 조용한 실내, 소음이 있는 실내, 조용한 실외 그리고 소음이 있는 실외 등 다양할 수 있다. 이러한 모든 환경에서도 안정적인 결과를 낼 수 있어야 한다는 점에서, 도전적인 과제임은 분명하다. 이러한 문제를 완벽히 풀어냈다고 볼 수는 없지만, 계속적으로 풀어나가고 있고, 음성 인식 결과만으로 문제를 해결해나가는 관점을 넘어, 자연어 이해 모델과 결부한 다양한 방법적인 고민을 통해 문제를 풀어 나가고 있다.

3-2. User Study

VUI의 수용도 관점으로 사용자 층을 Heavy User, Light User로 구분하고, 이들이 어떤 어려움을 겪고 그것에 반응 혹은 대응하는지 알아보기 위해 60명의 사용자를 대상으로 유저 스터디를 수행했다.

3-2-1. 참여자 (Participants)

VUI의 수용도 관점에서 사용자 층을 Heavy User 와 Light User로 구분하는 기준을 잡는 것은 본 연구에서 중요하다. VUI 참고 문헌의 기준을 활용해 [16] 참여자 군을 선별하기 위한 Pre-screening 설문을 준비했다. 참여자 군의 선별 기준은 다음과 같다. (표 1 참조)

표 1. Heavy User와 Light User 구분 기준

사용자 그룹	Frequency of Use	Recency of Use
Heavy User	월 3회 초과	최근 1주일 이내 사용

^⑨ <https://clova.ai/speech>

Light User	월 2회 미만	최근 1달 이내 미사용
------------	---------	--------------

위 기준을 만족하는 사용자 집단을 선별하기 위해서는 설문에 참여하는 사람이 조금 더 이해하기 쉬운 기준으로 풀어 설문 보기를 작성할 필요가 있었다. 실제 해당 설문 항목의 보기 기준은 다음과 같다.

- C1. 나는 어떤 형태의 VUI든 매일 사용하고 있다.
- C2. 나는 어떤 형태의 VUI든 1주일에 1~2회 이상 사용하고 있다.
- C3. 나는 어떤 형태의 VUI든 1달에 1~2회 가량 사용하고 있다.
- C4. 나는 어떤 형태의 VUI든 평균적으로 1달에 1번 이하로 사용하고 있다.
- C5. 나는 VUI를 사용해 보았으나, 언제 사용했는지 기억나지 않는다.
- C6. 나는 VUI를 사용한 경험이 없다.

위 6개의 카테고리로 사용자 군을 다음과 같이 나누게 되었다. (표 2 참조)

표 2. 기술 수용도에 따른 사용자 그룹핑

사용자 그룹	카테고리
Heavy	C1, C2
Intermediate	C3
Light	C4, C5, C6

위 표를 기준으로 Heavy User와 Light User에 해당하는 참여자 각 30명씩 총 60명의 참가자를 모집하였다. 각 사용자 그룹을 기준으로, 참가자의 성별과 나이에 대한 기초 통계 정보는 다음과 같다. (표 3 참조)

표 3. 참여자 정보

사용자 그룹	나이		성별	
	평균	표준편차	여	남
Heavy	34.27	6.41	11	19
Light	32.73	5.71	10	20

AiCall의 유저 스터디는 AiCall이 정식으로 릴리즈 되기 전, 사내 클로즈 베타 테스트 (Closed Beta Test, CBT)의 형태로 진행되었다 (2019년 11월). 본 CBT는 해당 제품의 출시 점검을 위한 것이기 때문에, 사내 보안 등의 문제로 외부 참여자를 모집하여 테스트를 진행하기 어려운 측면이 있었다. 그로 인해, 실제 참여자는 네이버 내부의 Clova 부문 그리고 Glace 부문의 임직원을 대상으로 진행되었다는 점을 밝힌다.

3-2-2. 진행 방식 (Procedure)

유저 스터디는 실험실 세팅으로 진행되지 않고, 실제 전화 사용 환경을 감안해 진행되었다. AiCall은 전화 사용을 기반으로 한 VUI이기 때문에, 참여자의 자연스러운 상호작용을 이끌어내기 위해서는, 그에 걸맞는 사용 환경을 제공할 필요가 있다고 판단했다. 그래서 사용자들이 실제로 원하는 시간에, 원하는 공간에서 참여할 수 있도록 테스트 환경을 구성할 필요가 있었으며, 참여 방식의 명확한 정의가 필요했다. 구체적인 테스트 참여 방식은 다음과 같다.

- 테스트는 하루를 기준으로 진행
- 테스트를 위한 전화번호는, 참여자 정보 수령 시 기입한 전화번호로만 진행
- 테스트 가능 시간은 미션 전달 후, 오후 8시 이전까지 진행
- 미션은 참여자의 회사 메일(@navercorp.com)로 평가 당일

오전 전달

- 미션은 총 4가지로 구성되어 있음 (표 4 참조)
- 참여자는 마지막 미션까지 끝낸 뒤, 가급적 바로 설문에 응답하고, 또 참여자 본인이 VUI와 상호작용하면서 겪었던 어려움을 자유롭게 기술
- 메일을 전달 받고, 진행에 어려움이 있는 경우, 평가 진행자에게 회사 메일 혹은 사내 메신저 (Line Works)를 통해 문의할 수 있음

참여자는 위 내용에 대한 전체적 가이드가 담긴 메일을 당일 오전 전달 받았으며, 참여자 중 일부는 해당 내용에 대한 문의를 메일과 사내 메신저를 통해 질문하였다. 질문에 대해서는, 평가를 방해하지 않는 선에서 답변을 제공했다. 참여자는 해당 테스트에 대한 참여가 완료된 경우, Glace CIC에서 준비한 소정의 네이버 포인트를 받았다^⑩.

표 4. 테스트 미션 상세

미션 Number	미션 내용
Mission 1	13명의 지인과 함께 아웃백 스테이크하우스 미금점을 방문합니다. 점심 시간에 예약이 가능한지 확인하고, 주차와 가는 길에 대해서 편하게 질문해주세요. (선호 시간: 12시)
Mission 2	남자친구 혹은 여자친구와 함께 기념일을 맞아 아웃백 스테이크하우스 미금점에 방문합니다. 점심 혹은 저녁 시간 중 특정 시간을 정해 예약을 진행해주세요. 원하시는 자리가 있는 경우, 자리에 대해서도 편하게 질문해주세요.
Mission 3	차를 가지고 지금 아웃백 스테이크하우스 미금점으로 가고 있습니다. 예약은 따로 진행하지 않았고,

^⑩ <https://pay.naver.com/about>

	부모님, 그리고 자녀 (혹은 조카들)와 함께 아웃백으로 가고 있는데, 정확한 가는 길과 부모님 및 어린이를 위한 편의시설 / 메뉴 등이 궁금한 상황입니다.
Mission 4	아웃백의 추천메뉴가 궁금합니다. 메뉴와 관련한 할인 사항도 함께 물어보세요. 어떠한 할인이 있는지, 각 할인 별 조건이 무엇인지도 확인해주세요.

미션은 AiCall의 전체적인 기능을 한 번씩 평가할 수 있는 흐름으로 구성되었다. 앞서 언급하였듯이, AiCall은 목적 지향형 대화가 진행되면서 동시에 자주 묻는 질문에도 답할 수 있도록 설계되어 있다. 또한, AiCall이 성공할 수 있는 영역 뿐 아니라, AiCall이 제공할 수 없는 정보 영역에 대한 내용 역시 위 미션에 포함되어 있다. 위 네 가지 미션을 통해 참여자는 AiCall의 기능을 전반적으로 탐색할 수 있다.

3-2-2-1. Speech User Interface Service Quality (SUISQ)

참여자들이 실제로 해당 VUI의 사용성을 어떻게 평가하는지 정량적으로 파악하기 위해 Speech User Interface Service Quality (SUISQ-MR) 설문을 활용하였다. 해당 설문은 VUI 사용성 중 전화 환경에서의 사용성을 평가하는 데 초점이 맞춰져 있다 [19, 21]. SUISQ 설문은 총 4가지의 상위 Factor와 (표 5 참조) 각 Factor 하위에 위치한 25개의 설문 항목으로 구성되어 있으며, 7점 리커트 척도를 (Likert Scale) 사용한다.

표 5. SUISQ의 4가지 Factor

Factors	Number of Items (개별 설문 항목 수)
User Goal Orientation	8개
Customer Service Behavior	8개
Speech Characteristics	5개

Verbosity	4개
-----------	----

SUISQ 설문은 각각 설문 항목 별 측정 목표의 유효성은 보장하면서 동시에, 항목 수를 줄인 두 가지의 압축 버전을 가지고 있다. 하나는 SUISQ-R (Reduced)이고, 다른 하나는 SUISQ-MR (Maximally Reduced)이다. SUISQ-R은 총 14개의 항목으로 구성되어 있으며, SUISQ-MR은 총 9개의 항목으로 구성되어 있다. 본 연구에서는 SUISQ-MR 설문을 사용했으며 (표 6 참조) 5점 리커트 척도를 활용했다.

표 6. SUISQ-MR (Maximally Reduced) 설문 항목

Factors	Questions
User Goal Orientation	I would be likely to use this system again. (이 시스템을 다시 사용하고 싶다.)
	I felt confident using this system. (이 시스템을 자신감 있게 사용할 수 있다.)
Customer Service Behavior	The system used everyday words. (이 시스템은 일상적인 단어를 사용해서 말한다.)
	The system seemed polite. (이 시스템은 공손하게 응대한다.)
Speech Characteristics	The system's voice sounded natural. (이 시스템의 목소리는 자연스럽다.)
	The system's voice sounded enthusiastic or full of energy. (이 시스템의 목소리에서 긍정적인 활력이 느껴진다.)
Verbosity	I felt like I had to wait too long for the system to stop talking so I could respond.

	(내가 말하려면 시스템의 말이 끝날 때까지 너무 오래 기다려야 한다.)
	The messages were repetitive. (이 시스템은 반복적으로 응답하는 경우가 많다.)
	The system was too talkative. (시스템이 너무 말을 많이 한다.)

3-2-2-2. VUI 사용 시 겪었던 어려움의 리포트 방식

평가가 실험실 환경에서 진행되지 않고, 원격으로 진행되었다는 측면에서, 사용자를 대상으로 인터뷰를 진행하는 것은 용이하지 않았다. 대신 참여자가 스스로 자신이 VUI를 사용하면서 느꼈던 어려움을 보고할 수 있도록 전체 설문 항목 내에 다음의 질문에 대한 주관식 답변을 받는 입력란을 포함하였다.

- 질문: AiCall의 총 4가지의 통화 미션을 완료하셨나요? 미션을 완료하셨다면, 사용하시면서 불편하거나 어려움을 겪었던 지점이 있다면 상세하게 기술해주세요.

위 질문에 대한 답변을 통해 VUI 사용자가 느끼는 어려움에 대해서 상세하게 이해하고 나아가 기술 수용 관점에서의 사용자 그룹 간의 어려움 대한 정도의 차이에 대해 이해할 수 있기를 기대했다. 해당 답변의 참여율과 적극성을 높이기 위해, 충실히 답변한 참여자에게 추가적인 리워드 (네이버 포인트)가 제공될 것으로 사전 안내되었다.

3-2-2-3. 통화 내용 자동 기록 (Call Session Viewer)

참여자들의 통화 기록은 그들의 동의 하에 연구 및 제품 개선의 목적으로 저장되었다. Clova의 Call Session Viewer는 진행된 통화의 내용을

자동으로 기록할 수 있도록 자체 개발되었다. 2020년 상반기 기준으로, 비즈니스 클라이언트를 포함하여, 외부에 공개되지 않은 제품이다. Call Session Viewer를 통해서 확인할 수 있는 내용은 다음과 같다 (표 7 참조).

표 7. Session Viewer 피쳐 리스트

Feature	설명
Session ID	세션 개별 부호이며, 모든 독립적인 콜에 대해 ID가 발급
Caller	전화를 건 번호
Respondent	전화를 받은 번호
Time	전화를 건 날짜 및 시간을 기록
Duration	통화 진행 시간
Call Detail	Caller 턴, Respondent 턴 모두가 발화 순서대로 기입. 사용자의 턴인 경우, 음성 인식된 결과가 텍스트로 표현
Play	각 턴의 실제 발화 내용이 녹음되어 있음

본 연구에서는 Caller 번호를 통해, 참여자가 실제로 참여를 진행했는지 확인하며, 몇 번의 통화를 진행했는지 확인할 수 있다. 또한, Call Detail의 Caller 턴의 텍스트로 음성 인식 결과를 확인할 수 있다. 이와 실제 발화를 비교해보려면 Play 버튼을 통해, 실제로 참여자가 발화한 내용이 ASR을 거쳐, 어떤 결과로 나오게 되었는지 알 수 있다. 실제 발화와 시스템에서 처리한 발화 사이의 차이를 알 수 있다는 점은, 향후 분석에 있어 유용하게 쓰일 수 있다.

3-3. 분석 방법 (Analysis Method)

본 연구의 유저 스테디는 Heavy User 그룹과 Light User 그룹으로 분류한 참여자 집단이 VUI를 사용하면서, 1) 어떤 어려움을 겪는지 그리고,

2) 그 어려움을 어떤 방식으로 대응 혹은 극복하는지 알아보기 위한 목적으로 수행되었다. 이를 위해서 유저 스터디를 진행하면서 확보된 정량, 정성 데이터의 분석 과정이 필요하다.

3-3-1. VUI 사용 시 겪는 어려움을 이해하기 위한 분석

먼저, 두 그룹의 참여자가 VUI 사용 시 겪는 어려움을 이해하기 위한 정량, 정성 데이터의 분석 과정을 거친다. 정량적 분석에 해당하는 것은 SUISQ-MR 설문 결과에 대한 스코어 분석이 있다. 정성적 분석은, 주관식 설문 형태로 참여자가 보고한 ‘사용 시 겪었던 어려움’에 대한 자료 분석이 있다.

3-3-1-1. 정량적 분석 방법

먼저 SUISQ-MR의 결과 분석을 통해 정량적으로 각 사용자 그룹의 VUI 사용성에 대해 분석하였다. SUISQ-MR에 대한 결과는 각 세부 문항 별로 평균을 집계하였고 (총 9개 문항). 상위 Factor로 한 번 더 평균에 대한 값을 구했다. 이 때 Factor는 앞서 언급하였던, 다음의 네 가지이다.

- User Goal Orientation
- Customer Service Behavior
- Speech Characteristics
- Verbosity

이렇게 구해진 스코어는 위 네 가지의 Factor를 대표하는 점수가 되며, 이를 통해 사용자가 해당 VUI의 사용성을 어떻게 평가하는지 알 수 있다.

본 연구의 목적이 상이한 두 사용자 그룹, 즉 Heavy User 그룹과 Light User 그룹을 비교하는 것이므로 연구 목적에 따라, SUISQ-MR 점수를 통해 실제로 두 사용자 그룹이 VUI 사용성을 평가한 것에 유의미한 차이가

있는지 파악하고자 했다. 서로 다른 두 그룹에 대해 평균의 유의미한 차이를 구하기 위해 Two-sample t-test를 수행하였다. 평균을 비교할 때의 단위는 세부 항목의 평균이 아닌, 상위의 네 가지 Factor에 대한 평균 스코어를 비교하는 것을 목표로 하였다.

3-3-1-2. 정성적 분석 방법

정성적 분석은 각 참여자 그룹이 리포트한 ‘VUI 사용 시 겪는 어려움’에 대한 정성 데이터에 대해 Grounded Theory에 [8] 근거하여 진행하였다. 단, 참여자는 연구원과의 대면 인터뷰 방식으로 사용 상의 어려움을 리포트하지 않고, 간접적인 방식인 설문을 통한 주관식 답변 형태로 스스로가 겪었던 어려움을 리포트하였다.

설문 결과는 Google Forms^⑪를 통해 종합하였으며, 참여자 그룹 별로 해당 결과를 나누어서 확인하였다. 해당 결과를 활용해, 우선적으로, 데이터에 대해 개방형 코딩을 하는 과정을 탐색적으로 진행했다. 이 때는, 기술 수용 정도에 따른 참여자 그룹 구分的 특수성에 대한 사전 지식 등을 투영하지 않고, 귀납적으로 발견되는 주요 키워드나 표현에 근거해 코딩을 진행했다. 해당 과정을 1차적으로 거치고 난 뒤, 2차로 추상화 작업을 진행했다.

나열된 코딩 결과물들은 몇 가지의 주요 개념으로 범주화될 수 있었다. 이렇게 범주화된 결과물은 참여자들이 리포트한 ‘VUI 사용 시 겪었던 어려움’의 개념으로 받아들여질 수 있으며, 해당 분석 결과물은 연구원의 추가적인 2차 해석이 포함되기 전, 참여자 관점에서의 어려움을 있는 그대로 판단할 수 있다는 점에서, 본 연구의 중심 결과로 활용되었다.

그 다음으로, 위 범주화된 결과를 활용해, VUI 기술과 디자인 관점으로 다시 한 번 주요 축을 기준으로 개념들을 추상화시키는 작업을 진행했다.

^⑪ <https://www.google.com/forms/about/>

즉, 관련된 어려움이 내포하는 기술적인, 혹은 디자인 관점의 문제는 무엇일지를 알아내기 위한 작업이었다. 이렇게 도출된 VUI 디자인 및 개발과 관련된 주요 개념들은 향후, 참여자들이 VUI 상에서 겪게 되는 어려움에 대응하거나, 극복하는 방법을 대화 분석을 통해 알아보는 과정에서 활용된다.

3-3-2. VUI 사용 시 겪는 어려움에 대한 대응 방식 분석

두 참여자 그룹이 어려움에 대응하는 방식을 알아보기 위해, Call Session Viewer에 저장된 대화 흐름과, 통화 녹음본을 활용하여 분석이 진행되었다.

3-3-2-1. 대화 흐름 분석

참여자가 AiCall과 어떻게 대화를 나누었는지 알아보기 위해, 통화 별로 대화의 흐름을 분석한다. 먼저 Call Session Viewer에 기록된 텍스트를 통해, 해당 대화의 흐름을 이해할 수 있다. 이 때, Call Session Viewer는 사용자의 발화에 대해 음성 인식된 결과를 텍스트로 보여주기 때문에, 해당 텍스트는 실제 유저가 진행한 발화와 다를 수 있다. 예시는 다음과 같다 (표 8 참조).

표 8. 실제 발화와 Session Viewer 결과 비교 (참여자 통화 중 발체)
(* U: User, S: System)

실제 발화		Session Viewer 결과	
U	영업시간 언제부터인지 궁금해요	U	영업시간 언제 들어온 직후 코미
S	쉬는 날 없이 매일 오전 11시 반부터 오후 10시까지 영업해요.	S	쉬는 날 없이 매일 오전 11시 반부터 오후 10시까지 영업해요.
U	매장 정확한 주소가 어떻게 되나요	U	매장 정확한 주소가 어떻게 되나요

S	경기도 성남시 분당구 돌마로에 있어요. 정확한 주소를 문자로 보내드릴까요?
U	네네
S	문자로 보내드렸습니다. 혹시 더 필요하신 거 있으신가요?

S	경기도 성남시 분당구 돌마로에 있어요. 정확한 주소를 문자로 보내드릴까요?
U	쉬 내니
S	문자로 보내드렸습니다. 혹시 더 필요하신 거 있으신가요?

위 표를 통해 볼 때 우측 표의 Session Viewer 결과 중 User 발화는 왼쪽의 실제 발화와 다른 것을 알 수 있다. Session Viewer 결과에서 실제 발화를 확인하기 위해서는, Session Viewer에서 제공하는 발화 음성 녹음을 활용한다. 이를 통해, 실제로 참여자가 말한 발화의 음성 인식 결과, 그리고 그것이 어떻게 자연어 이해를 거쳐 대화 운영 모듈에서 매핑된 답변을 출력하게 되었는지 일련의 과정을 이해할 수 있게 된다.

또한, Session Viewer를 통해 사용자의 발화 인식 결과 그리고 대화 흐름 양상의 특이성을 확인하는 과정을 통해, 사용자가 문제에 직면했을 때, 이를 어떻게 풀어나가는지 파악할 수 있다. 각 참여자가 Self-report 방식으로 제기한 문제에 대해서, 어떤 형태로 문제를 해결해 나가는지 파악하는 과정을 통해 참여자 그룹 별로 어려움을 극복하는 방식의 차이를 비교할 수 있다.

4. 연구 결과 (Result)

본 장에서는 유저 스터디를 통해 도출된 정성적, 정량적 분석 결과를 제시한다. 크게 연구 결과에 대해서는 두 가지 단계를 거쳐 논의 주제를 제시한다. 첫 번째 단계는 두 가지의 참여자 그룹이 VUI를 사용하면서 겪는 어려움의 형태에 대한 탐색이다. 이 과정에서는, SUIQ-MR 점수에 대한 Two sample t-test를 통해, 두 그룹 간에 VUI 사용성에 있어, 유의한 차이가 있는 항목이 무엇인지 확인한다. 또한 각 사용자 그룹이 제공한 Self-report 형태의 정성적인 자료를 분석하고 그 결과를 제시한다. 두 번째 단계에서는 정의된 어려움의 형태에 대해, 두 참여자 그룹이 어떻게 대응하거나 극복하는지 확인한다. 이 단계에서는 Call Session Viewer의 참여자 Call log (텍스트 및 통화 녹음) 데이터를 분석하여 결과를 도출한다. 위 과정을 통해, 토론 주제를 제시하고, 상이한 사용자 그룹을 위한 맞춤형 VUI 제공을 위해 고려할 사항을 논의한다.

4-1. SUIQ-MR 결과 분석

참여자가 생각하는 VUI 사용성에 대해 알아보기 위해, SUIQ-MR 설문 결과에 대한 정량적인 분석을 진행했다. 정량적인 분석 결과는 다음과 같다 (표 9 참조).

표 9. SUIQ-MR 결과에 대한 Two-sample t-test

Factors	사용자 그룹	평균의 차 ^⑫ (H-L)	표준 편차	t	p
User Goal Orientation	Heavy	0.17	0.83	0.785	0.436
	Light		0.82		

^⑫ 대외비 이슈로 인해 각 평균값을 제공하지는 못함에 양해를 구함.

Customer Service Behavior	Heavy	-0.03	0.70	-0.176	0.861
	Light		0.76		
Speech Characteristics	Heavy	0.19	0.85	0.723	0.473
	Light		1.10		
Verbosity	Heavy	0.58	0.81	3.000	0.004**
	Light		0.68		

SUISQ-MR 결과의 상위 4가지의 Factor를 기준으로 Two-sample t-test를 수행한 결과, Verbosity, 즉 사용자가 느끼는 대화형 에이전트의 발화량 및 반복 발화의 과다는 두 그룹 간 유의미한 차이가 있는 것으로 나타났다. 특히 Heavy User 그룹이 Light User 그룹보다, 해당 항목에 대해 더 많이 문제 삼았음을 확인할 수 있는데, 평균 차가 0.58점으로 나머지 세 가지 항목에 비해 2배 이상 크다는 것을 확인할 수 있다. 참고로, 세부 항목 간의 평균 차이는 다음과 같다 (표 10 참조).

표 10. SUISQ-MR 세부 항목에 대한 평균 차

Factors	세부 항목	평균의 차 (H-L)
User Goal Orientation	I would be likely to use this system again. (이 시스템을 다시 사용하고 싶다.)	0.03
	I felt confident using this system. (이 시스템을 자신감 있게 사용할 수 있다.)	0.30
Customer Service Behavior	The system used everyday words. (이 시스템은 일상적인 단어를 사용해서 말한다.)	0.00
	The system seemed polite. (이 시스템은 공손하게 응대한다.)	-0.07

Speech Characteristics	The system's voice sounded natural. (이 시스템의 목소리는 자연스럽다.)	0.03
	The system's voice sounded enthusiastic or full of energy. (이 시스템의 목소리에서 긍정적인 활력이 느껴진다.)	0.33
Verbosity	I felt like I had to wait too long for the system to stop talking so I could respond. (내가 말하려면 시스템의 말이 끝날 때까지 너무 오래 기다려야 한다.)	0.77
	The messages were repetitive. (이 시스템은 반복적으로 응답하는 경우가 많다.)	0.23
	The system was too talkative. (시스템이 너무 말을 많이 한다.)	0.73

4-2. VUI 이용 시 겪는 어려움의 유형 (Difficulty Types)

유저 스터디의 결과를 통해 Heavy User 그룹에 속한 참여자와 Light User 그룹에 속한 참여자들이 어떤 유형의 어려움을 경험했는지 파악했다. 어려움의 유형은 최종적인 코딩 과정을 거쳐, 15가지가 도출되었다 (표 11 참조). 위 15가지의 범주화된 어려움 유형에 대해, Heavy User 그룹과 Light User 그룹 도합 총 133건이 리포트 되었다. 이 133건은 1차 코딩 결과에 대한 2차 범주화 작업이 완료된 건들에 대해 수치를 센 것임을 밝혀둔다. 참고로, Heavy User의 리포트 수가 Light User의 리포트 수 대비, 약 2배 가량 많았다¹³⁾.

¹³⁾ 대외비 이슈로 인해 각 사용자 그룹 별 정확한 리포트 건수는 밝히지 못함에 양해를 구함.

표 11. 15가지 VUI 사용 시 경험한 어려움의 유형¹⁴⁾

Type	Difficulty	Heavy	Light
D1	에이전트가 같은 말을 너무 반복적으로 함	0.21	0.05
D2	에이전트가 내가 했던 말을 정확히는 이해하지 못한다고 생각됨	0.19	0.09
D3	시스템이 내 말을 인식하는 것에 문제가 있다고 생각됨	0.14	0.13
D4	에이전트의 목소리가 자연스럽지 않음	0.10	0.03
D5	에이전트가 내가 말하고 있는 도중 내 말을 끊음	0.07	0.04
D6	가끔 지금 무슨 말을 해야할 지 모르겠는 경우가 있음	0.05	0.05
D7	시스템이 내 요청을 처리하는 논리적 흐름이 이상하다고 생각됨	0.04	0.00
D8	시스템이 멀티턴 대화를 지원하지 않는 것으로 보임	0.04	0.01
D9	에이전트가 하는 말을 내가 끊을 수 없음	0.04	0.00
D10	에이전트가 한 번 말을 하면 너무 길게 함	0.03	0.01
D11	에이전트가 제공하는 정보의 양이 불충분함	0.02	0.01
D12	대화 흐름이 어색하다고 생각됨	0.02	0.01
D13	에이전트의 발화 속도가 너무 빠름	0.01	0.00
D14	에이전트가 한 번에 너무 다양한 종류의 정보를 내게 요청함	0.01	0.00
D15	에이전트의 응답 속도가 느린 것 같음	0.01	0.01
		1.00	0.46

¹⁴⁾ 해당 표의 결과는 대외비 이슈로 인해 전체 총합 133건에 대한 상대적 수치를 기입하였으니, 양해를 바랍.

특징적으로, Heavy User 그룹이 보고한 어려움의 종류가 Light User 그룹이 보고한 종류를 모두 아우른다. 실제로, 결과를 정리해보면, Light User 그룹이 보고한 어려움의 종류는 총 11개로 정리되었음을 확인할 수 있다. 다음 보여지는 어려움의 종류는 Light User 그룹에서 발견되지 않은 형태이다.

- D7: 시스템이 내 요청을 처리하는 논리적 흐름이 이상하다고 생각됨
- D9: 에이전트가 하는 말을 내가 끊을 수 없음
- D13: 에이전트의 발화 속도가 너무 빠름
- D14: 에이전트가 한 번에 너무 다양한 종류의 정보를 내게 요청함

다음으로 위 어려움 유형에 대해 그룹 별로 그 비율을 확인해 보았다. 위 비율을 확인함으로써 각 그룹 별로 문제 삼는 어려움의 유형에 대한 순위를 파악할 수 있다 (그림 3 참조).

그림 3을 통해 파악할 수 있는 점은, Heavy User 그룹과 Light User 그룹 내에 어려움 유형의 순위 구성에 차이가 있다는 점이다. 흥미로운 점은 Top 3 어려움 유형인 D1 (에이전트가 같은 말을 너무 반복적으로 함)과 D2 (에이전트가 내가 했던 말을 정확히는 이해하지 못한다고 생각됨) 그리고 D3 (시스템이 내 말을 인식하는 것에 문제가 있다고 생각됨)의 각 그룹 내 순위 구성이 반대로 되어 있다는 점이다. 즉, Heavy User 그룹 내에서는 D1 - D2 - D3 순으로 1위부터 3위가 구성되어 있다면, Light User 그룹 내에서는 D3 - D2 - D1의 순서로 1위부터 3위가 구성되어 있다.

이는, 4위부터 6위에 해당하는 어려움 유형에도 마찬가지로 적용된다. D4 (에이전트의 목소리가 자연스럽지 않음), D5 (에이전트가 내가 말하고 있는 도중 내 말을 끊음) 그리고 D6 (가끔 지금 무슨 말을 해야할 지

모르겠는 경우가 있음)에 대해 Heavy User 그룹은 D4 - D5 - D6 순으로 4위부터 6위가 구성되어 있다면, Light User 그룹 내에서는 D6 - D5 - D4 순으로 4위부터 6위가 구성되어 있다. 이는 두 그룹이 생각하는 어려움 유형의 순위가 서로 다를 수 있음을 나타낸다.

참고로, 본 연구원은 각 사용자 그룹의 Self-report 데이터를 대상으로, 어려움 유형을 도출하기 위해 1차적으로 개방형 코딩을 진행했다. 15가지의 어려움 유형을 범주화 하기 전 중복을 제외한 1차 코딩 결과물은 37건이고, 중복을 포함한 경우 전체 참여자 기준 리포트 된 결과물은 188건이다.

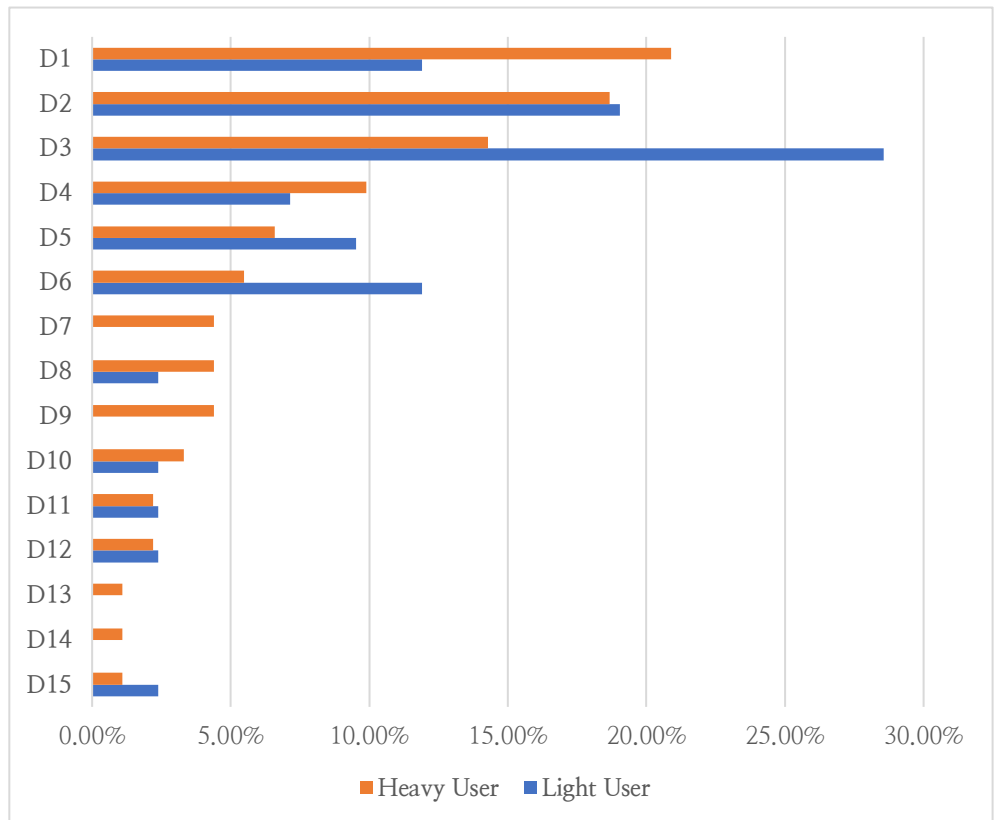


그림 3. 각 사용자 그룹 별, 어려움 유형 비율^⑮

^⑮ 대외비 이슈로 인해 Percentage 비율로 표현함에 양해를 구함.

4-2-1. Heavy User 그룹의 어려움 유형

Heavy User가 VUI를 사용하면서 경험하고 리포트한 어려움의 유형은 Light User보다 많다. 그들이 Self-report 형태의 주관식 설문에서 어려움을 어떻게 표현했는지 역시 중요한 자료가 될 수 있으므로, 그 표현 방식에 대한 결과를 제시할 필요가 있다. 먼저, 빈도 수가 가장 높았던 D1 (에이전트가 같은 말을 너무 반복적으로 함)에 대해 Heavy User 그룹의 참여자들은 해당하는 어려움 유형을 다음과 같이 표현했다.

“매번 시간이나 인원을 확인하는 것보다 최종적으로 한 번만 확인하면 좋겠습니다. 이 부분이 좀 거슬리네요. 예를 들어 시간과 인원을 한 번 예약하면, 계속해서 ‘몇 월, 몇 일, 언제, 몇 명으로’라는 멘트가 고정적으로 계속 나오는데요, 이 부분이 사용을 진행하는데 꽤나 불편하고 진행에 직접적인 방해가 된다고 생각되네요.” – P21 Self-report 에서 발췌

“반복적으로 점원 연결로 보내주는 것이 부담스러움. 정말 복잡하고 어려운 주문을 하게 될 때는 애초에 점원 연결을 할 것 같음. AI가 ‘점원 연결해드릴까요’라고 말하는 순간 엄청 하기 싫어짐.” – P12 Self-report에서 발췌

Heavy User 그룹의 참여자들은 명확한 확인을 위한 프로세스, 그리고 문제 해결을 위한 점원 연결 발화 등에 대해 스스로 문제를 풀어가는데 방해되는 요소이며, 이러한 것이 VUI를 사용하는 데 겪게 되는 어려움이라고 표현했다.

또한, 특징적으로 Light User 그룹의 어려움 유형에는 나타나지 않았지만, Heavy User 그룹의 어려움 유형에서만 보고된 D9 (에이전트가 하는 말을 내가 끊을 수 없음)에 대해서 그들은 이렇게 표현했다.

“예약 내용 설명이 명확한 건 좋은데, 중간에 말을 끊을 수 있어야 할

것 같습니다. 중간에 말을 끊을 수가 없었어요. 무시하고 자기 말만 하는 것 같아 진행이 좀 어렵네요. 화도 나요.” - P11 Self-report에서 발췌

“테스터들은 환경을 알고 있기 때문에 AiCall의 문장이 끝나기를 기다렸다가 질문을 합니다. 그렇기 때문에 성공률이 높은데요. 조금만 말이 끝나기 전에 들어가면 이해하지 못하는 경우가 발생하여 불편합니다. (사용자 발화가) 겹쳐진 부분에 대해 말을 못 듣는 걸까요?” - P2 Self-report에서 발췌

이들은 시스템의 발화가 끝나기 전에 말을 하는 경우에 대해, 적어도 말을 끊을 수 있거나 (Barge-in), 아니면 중간에 끊지 못한다면, 말이 겹쳐진 부분 (overlapped)이 처리될 수 있어야 한다고 지적했다. P11은 이렇게 말이 끊어지지 않고, 시스템이 발화를 이어나가는 상황에 대해 ‘무시하고 자기 말만 하는 것 같아 진행이 좀 어렵고, 화도 난다’라고 감정적인 표현을 하기도 했다.

4-2-2. Light User 그룹의 어려움 유형

Light User가 VUI를 사용하면서 경험하고 리포트한 어려움의 유형은 Heavy User의 그것보다 적지만, 흥미롭게도 Light User 그룹의 어려움 유형 순위는 Heavy User 그룹의 순위와 다르다. Light User 그룹이 가장 많이 리포트한 어려움 유형은 D3 (시스템이 내 말을 인식하는 것에 문제가 있다고 생각됨)로, 자신의 말을 시스템이 잘 알아듣는지 확신하지 못한다고 생각하는 것과 관련이 있다.

“날짜, 시간, 인원 수 정보를 한 번에 정확히 인식할 수 있으면 좋겠습니다.” - P47 Self-report에서 발췌

“제 발음 때문인지는 모르겠지만, 숫자 등에 대한 인식률이 떨어지는 거 같네요. 이번 주 28일이라는 말을 계속 못 알아 듣습니다. (그래서 음성

비서를 잘 안씀)” – P44 Self-report에서 발췌

Light User 그룹은 대화형 에이전트가 자신의 말의 뜻을 ‘이해하지’ 못하는 것에 앞서, 자신의 말을 제대로 ‘듣지’ 못하는 것에 대해서 어려움을 느끼고 있음을 위와 같이 표현했다. 실제로, 에이전트의 성능이 아닌, 참여자 자신의 문제일 수 있다고 표현하면서, 이것이 참여자 본인이 VUI를 쓰지 않는 이유와 연결하는 형태의 발언을 한 경우도 보인다.

또한, 어려움 유형 D6 (가끔 지금 무슨 말을 해야할 지 모르겠는 경우가 있음) 역시, Light User 그룹에서 상대적으로 높게 나오는 것을 확인할 수 있다. 이는 VUI의 해석 가능성과도 연결된 문제라고 볼 수 있는데, 다음과 같이 해당 어려움을 표현했다.

“런치타임은 몇시부터 몇시예요, 매장은 어디 어디 앞 쪽에 있어요. 하고 ‘혹시 더 궁금하신 게 있나요?’ 멘트가 나오지 않아 전화가 끊어진 것인지 무엇인지 잘 모르겠던 상황이 있었습니다. (버그일 수도 있겠네요...)”
– P42 Self-report에서 발췌

“음성 비서 형태를 처음 쓰는 것은 아니지만, 전화 시작 시에 AI 응대에 대한 안내를 사용자에게 고지하여 주면 보다 효과적인 통화가 될 수 있을 것 같습니다.” – P56 Self-report에서 발췌

D6 어려움 유형의 비율은 Light User 그룹에서 상대적으로 높다. 어떻게 제품을 이용하면 좋을지, 전반적으로 혹은 부분적으로 VUI가 안내하고 이끌어 줄 필요가 있다는 점을 리포트했다.

4-3. VUI 구성 단위로 재범주화된 어려움 유형

사용자 관점의 어려움 유형은 시스템 관점으로 파악될 수 있다. 즉, VUI를 구성하는 기술 요소의 어떤 측면이 사용자로 하여금 어려움을

유발하였는지 유형화 또는 범주화 시켜볼 수 있다는 점이다. 본 연구원은, 도출된 15가지의 어려움 유형을 각 기술 단위 및 디자인 요소와 연결 지어 시스템 관점으로 바라보았다. 그 결과는 다음과 같다 (표 12 참조).

표 12. 어려움 유형과 관련한 VUI 구성 단위 (기술 및 디자인)¹⁶

Type	VUI 구성 단위				
	ASR	Data & NLU	DM	TTS	Conversation Design
D1			V		V
D2	V	V			
D3	V	V			
D4				V	
D5	V				
D6					V
D7			V		
D8		V			
D9			V		
D10			V		V
D11		V			V
D12	V	V	V		V
D13				V	
D14					V
D15	V		V		

위 범주화 기준을 활용하여, 어려움 유형에 영향을 미치는 VUI 구성 단위에 대한 비율을 Heavy User와 Light User 전체에 대해 확인을 하면 다음과 같다.

¹⁶ 어려움 유형 별 가중치에 대해서는 대외비 문제로 인해 본 표에서 제시하지 않음에 대하여 양해를 구함.

표 13. 어려움 유형과 관련된 VUI 구성 단위에 대한 사용자 그룹별 비율¹⁷

VUI 구성 단위	Heavy	Light
ASR	0.26	0.17
Data & NLU	0.25	0.14
DM	0.22	0.05
Conversation Design	0.21	0.09
TTS	0.07	0.02
	1.00	0.47

Heavy User, Light User 그룹 모두 어려움 유형과 관련된 VUI 구성 단위의 Top1, 2가 음성 인식 (ASR) 그리고 데이터와 자연어 이해 (Data & NLU)로 확인된다. 물론, Heavy User 그룹은 Top1, 2의 비율 차이가 크지 않은 것에 비해, Light User 그룹은 그 순위에 대한 비율 차이가 Heavy User의 그것보다 상대적으로 크다는 것을 알 수 있다. Top3, 4에 대한 순위는 양쪽 그룹 간 차이가 있는데, Heavy User 그룹은 근소한 차이로 대화 운영 (DM), 대화 디자인 (Conversation Design) 순으로 순위가 구성된다. 하지만 Light User 그룹은 상대적으로 대화 디자인이 Top3에 위치하며 그 뒤로 대화 운영이 순위를 차지하고 있다. 모든 그룹에서 음성 합성 (TTS)과 관련된 어려움은 낮은 비중을 차지하는 것으로 파악된다.

어려움 유형을 시스템 관점으로 범주화하게 되면, 다음과 같이 사용자 그룹이 어려움을 느끼는 데에 영향을 주고 있는 VUI 구성 단위에 대해 더 쉽게 알 수 있으며, 이는 논의 주제로 활용될 수 있다.

4-4. VUI 사용 시 맞이하는 어려움을 대응하는 방식

본 연구원은, Call Session Viewer를 통해 실제로 사용자들이 VUI 사용 시 어려움을 맞닥뜨리는 양상 혹은 구체적인 원인과, 그에 대한 반응 혹은

¹⁷ 어려움 유형에 대해 보고한 빈도를 가중치로 활용하여 적용함.

대응 과정을 관찰하였다. 이에 대해 Heavy User 그룹과 Light User 그룹은 그 대응 과정이 어떠한지 확인할 수 있다. 본 연구에서는 모든 어려움 유형에 대해 다루지 않고, 특징적인 어려움 유형 몇 가지에 대해 Heavy User와 Light User 간 어려움 유형 대응 방안이 어떠한지를 확인하고자 했다.

4-4-1. 어려움 유형 D1에 대한 대응 방식

어려움 유형 D1 (에이전트가 같은 말을 너무 반복적으로 함)에 대한 Call log를 탐색적으로 확인하고, 그에 대해 Heavy User 그룹과 Light User 그룹이 어떻게 반응 혹은 대응하는지 살펴보았다. 먼저 Heavy User 그룹의 반응 혹은 대응하는 방식에는 아래의 3가지 형태가 있었다.

- 평가 Task와 관련 없는 부정적 감정 표현
- 명시적인 통화 종료 의사 표현
- 잘못된 진행 중임에도, 진행 중인 대화를 마무리하고 (Sequence 종료¹⁸), 맥락을 파악해 나가며 대화 수정 (Repair) 과정 진행

먼저 D1 어려움 유형에 대해 Heavy User 그룹은 평가 Task와 관련 없는 부정적 감정을 표현하는 경우가 있었다. Call log의 탐색적 확인 결과, Heavy User 그룹에서 특징적으로 부정적 감정 표현이 많았는데, D1 어려움 유형과 관련되어서 많이 발견되었음을 확인할 수 있었다. 실제 사례 중 하나를 발췌하여 보면 다음과 같다 (표 14 참조).

표 14. D1 유형에 대한 부정적 감정 표현 사례

실제 발화			Session Viewer 결과	
1	S	단체 예약은 매장 직원 분께 확인을 해야 되거든요. 연결해 드릴까요?	S	단체 예약은 매장 직원 분께 확인을 해야 되거든요. 연결해 드릴까요?

¹⁸ 소단위 대화 종료를 의미 (Conversation Analysis 학문 분야의 용어 차용)

2	U	아뇨 9명도 단체예요?	U	아뇨 9명도 단체요
3	S	단체 예약은 매장 직원 분께 확인을 해야 되거든요. 연결해 드릴까요?	S	단체 예약은 매장 직원 분께 확인을 해야 되거든요. 연결해 드릴까요?
4	U	아뇨 아뇨 아뇨. 몇 명부터 단체예요?	U	아니 아니 몇명부터 단체예요
5	S	자세한 예약정보는 매장 직원 분께 문의하셔야 돼요. 연결해 드릴까요?	S	자세한 예약정보는 매장 직원 분께 문의하셔야 돼요. 연결해 드릴까요?
6	U	아니요. 근데 말투가 왜 그래요?	U	아뇨 근데 말투가 왜 그래요

위 Call log를 확인하였을 때, Turn 1, 3에서 에이전트가 사용자의 의도를 정확히 파악하지 못하고 반복적인 답변을 한 것을 확인할 수 있다. 이에 대해 사용자는 Turn 4에서 진행 중인 대화 sequence를 확장하여, 단체 인원 에 대한 정보를 문의했으나, Turn 5에서 에이전트는 사용자의 의도를 통화 내에서 처리하지 못한다고 판단하여, 외부 직원으로 넘기려고 한다. Turn 6는 반복적으로 사용자의 의도를 파악하지 못한 것을 포함한 문제 해결이 안되는 상황을 인지하게 된 사용자가 특정 액션을 할 것이라 예상할 수 있는 Turn이다. 이 때 사용자는 실험 태스크 진행과 관련 없는, 에이전트 말투에 대한 지적, 더 추상적으로 표현하자면, 위 과정에 대한 부정적 감정을 표현하는 것을 확인할 수 있다.

한편, 반복적인 발화로 불편함 혹은 대화가 진행되지 않음을 판단했을 때, 명시적으로 통화 종료 의사를 표현하는 형태의 대응을 하는 경우도 보였다 (표 15 참조).

표 15. D1 유형에 대한 명시적인 대화 종료 의사 표현 사례

실제 발화	Session Viewer 결과
-------	-------------------

1	U	부스석이 있어요? 부스석.	U	부 스 석 있어요 부스 성
2	S	죄송한데, 오시는 날짜랑 시간. 인원 다시 한 번 말씀해주실 수 있을까요?	S	죄송한데, 오시는 날짜랑 시간. 인원 다시 한 번 말씀해주실 수 있을까요?
3	U	예약말고여. 부스석이 있어요?	U	예약말고 여부 스 석이 있어요
4	S	죄송한데 다시 한 번 말씀해주시겠어요?	S	죄송한데 다시 한 번 말씀해주시겠어요?
5	U	전화 끊어주세요.	U	하나 빼주세요

위 Call log를 확인해보면, 사용자는 Turn 2를 통해 Turn 1의 사용자 발화를 에이전트가 이해하지 못했음을 확인했다. 그래서 Turn 2에서 에이전트가 했던 말의 카테고리인 ‘예약’과 관련한 발화가 아님을 명시적으로 Turn 3에서 표현하고 (“예약말고여.”), 다시 대화 수정 (Repair)의 과정을 거친다 (“부스석이 있어요?”). 하지만, 이러한 명시적인 대화 수정 (Repair) 과정에도 불구하고 에이전트가 Turn 4에서 다시 Turn 2의 연장선상인 알아듣지 못했다는 발화를 하는 것을 확인하고, 사용자는 Turn 5에서 통화 종료 의사를 명시적으로 표현한다. 물론, 이 발화 역시 인식이 잘못된 이유로 이후 발화가 통화 종료로 연결되지는 않았지만, 이렇게 명시적으로 대화를 종료하겠다는 의사, 즉 지금의 대화 흐름이 잘못되었을 때 에이전트로 하여금 잘못된 상황에서 벗어나겠다는 형태의 대응을 명시적으로 표현하는 것은 Heavy User 그룹의 다른 사례에서도 발견되었다.

위 형태와 유사한 맥락으로, 지금 진행 중인 대화를 마무리하고, 다시 대화 수정 (Repair)이 가능한 상황을 판단하기 위해 지켜보는, 맥락 파악형 대화 수정 (Repair) 진행을 이끌어 나가는 사용자 또한 존재했다. 다음의 대화는 그 사례를 대표한다 (표 16 참조).

표 16. D1 유형에 대한 맥락 파악형 Repair 사례

실제 발화			Session Viewer 결과	
1	U	아, 예. 알겠습니다. 언제까지 하나요 참, 토요일은?	U	아예 알겠습니다 언제까지 안에 잠 들은
2	S	고객님, 아기 의자 필요하세요?	S	고객님, 아기 의자 필요하세요?
3	U	아기 의자.... 예. 준비해주세요.	U	예약이 의장이 계신 대 주세요
4	S	고객님, 아기 의자 필요하세요?	S	고객님, 아기 의자 필요하세요?
5	U	예. 준비해주세요.	U	예 준비해주세요
6	S	혹시 의자 몇 개 필요하신가요?	S	혹시 의자 몇 개 필요하신가요?
7	U	의자 1개요.	U	의자 1개
8	S	아기 의자 1개 준비해드리겠습니다. 전화주신 ****번으로, **월 **일 토요일 오후 두시에, 세 분으로 예약해드렸습니다. 혹시 더 필요하신 거 있으신가요?	S	아기 의자 1개 준비해드리겠습니다. 전화주신 ****번으로, **월 **일 토요일 오후 두시에, 세 분으로 예약해드렸습니다. 혹시 더 필요하신 거 있으신가요?
9	U	언제까지 영업하나요?	U	언제까지 영업하나요
10	S	주문은 9시까지 받고 있구요, 저녁 10시에 문을 닫아요.	S	주문은 9시까지 받고 있구요, 저녁 10시에 문을 닫아요.

위 대화 내용을 파악해보면, 사용자는 Turn 1에서 문의한 내용을 Turn 10에서 들을 수 있었다. 하지만, 사용자 스스로가 대화 맥락에 대해서

이해를 하면서, 현재 시스템이 지원하지 못하는 상황을 이해하고, 7 turn을 희생하는 모습을 확인할 수 있다. 에이전트의 아기 의자 확인 대화 sequence를 종료한 뒤에서야, Turn 9에서 다시 질문을 진행해 원하는 답변을 얻어냈다. 앞선 부정적 감정 표현의 형태, 그리고 명시적인 통화 종료 의사 표현과는 상반되는 모습으로 보여질 수 있으나, 이는 대화 시스템의 흐름에 대한 이해를 기반으로 한, 또한 원하는 답을 어려움 없이 얻기 위한 효과적인 대화 수정 (Repair) 전략으로 보여진다.

다음으로, Light User 그룹은 어려움 유형 D1에 대해 어떻게 반응 혹은 대응하는지 살펴보았다. Light User 그룹에 있어 D1은 3순위에 올라와 있는 어려움 유형이라, Heavy User 그룹에 비해 문제 심각성은 크지 않을 수 있다. 하지만, 그 반응 혹은 대응 방식이 어떠한지는 알아볼 필요가 있다. 크게 다음과 같은 유형으로 D1에 대해 대응하는 경향성을 보인다.

- 반복 발화에 대해 대화 수정 (Repair) 지속 시도
- 맥락과 관계 없는 에이전트 발화에 의존하는 경향성

먼저, 반복 발화에 대한 대화 수정 (Repair)을 지속적으로 시도하는 경향성을 확인할 수 있다. 아래의 사례는 그 양상이 드러나는 대화 중 일부이다 (표 17 참조).

표 17. D1 유형에 대한 지속 Repair 사례

실제 발화			Session Viewer 결과	
1	S	준비해 드릴게요. 혹시 몇 시에 오시나요?	S	준비해 드릴게요. 혹시 몇 시에 오시나요?
2	U	한 3시에 갈 거 같아요.	U	가는 3시 쯤 같아요
3	S	당일 예약은 3시간 이후 좌석부터 가능하세요. 예약 가능한 시간으로 다시 한 번	S	당일 예약은 3시간 이후 좌석부터 가능하세요. 예약 가능한 시간으로 다시 한 번

		말씀해주시겠어요?		말씀해주시겠어요?
4	U	한 5시에 갈 거 같아요.	U	한 5시에 갈 거 같아요
5	S	당일 예약은 3시간 이후 좌석부터 가능하세요. 예약 가능한 시간으로 다시 한 번 말씀해 주시겠어요?	S	당일 예약은 3시간 이후 좌석부터 가능하세요. 예약 가능한 시간으로 다시 한 번 말씀해 주시겠어요?
6	U	어…。 그러면, 저녁 7시에 가능할 거 같아요.	U	그러면 저녁 7시에 가능할까요

위 사례를 보면, Turn 3으로부터 Turn 2의 사용자 발화를 에이전트가 이해하지 못했음을 사용자가 파악했음에도 불구하고, Turn 4, 6에 걸쳐서 반복적으로 대화 수정 (Repair)을 지속하는 경향성을 파악할 수 있다. 실제로 Turn 3, 5를 통해 지속적으로 반복된 발화가 나오는 것은 문제 상황이 될 수 있음에도 불구하고, 사용자는 이에 대해 특별한 문제 제기 없이, Turn 6에서도 지속적으로 에이전트의 요청에 맞춰서 대화 수정 (Repair) 발화를 진행하는 모습을 보였다. 이는, Heavy User 그룹에서 반복 발화가 나왔을 때 보여지는 반응 혹은 대응과는 다른 양상이며, 논의 주제에서 다뤄질 예정이지만, 해당 상황에 대한 문제 심각도 인식과 연관이 있을 수 있다고 보여진다.

또한, Light User 그룹의 대화에서는, 반복 발화의 문제가 발생하는 지점에서, 맥락과 관계 없는 에이전트의 발화에 의존하는 경향성을 보였다 (표 18 참조).

표 18. D1 유형과 관련 없는 맥락에 의존하는 사례

	실제 발화		Session Viewer 결과	
1	S	고객님 지금 예약 도와드릴까요?	S	고객님 지금 예약 도와드릴까요?
2	U	어…。 아뇨. 제가 나중에	U	아녜요 제가 나중에

		전화할게요.		전화할게요
3	S	그러면 예약 도와드릴까요?	S	그러면 예약 도와드릴까요?
4	U	제가 나중에 전화드리겠습니다.	U	제가 나중에 전화드리겠습니다
5	S	혹시 몇 분이서 오신다고 하셨죠?	S	혹시 몇 분이서 오신다고 하셨죠?
6	U	13명이에요.	U	13명이에요

사용자는 Turn 2, 4에서 비교적 명시적으로, 물론 Heavy User 그룹의 사례와 같이 직접적이지는 않지만, 통화 종료를 요청한다. 하지만, Turn 5에서 에이전트가 다시 대화를 리드하는 형태로, 예약 관련 내용을 진행하기 위한 방문 인원 수를 문의하자, 사용자는 ‘13명이에요’라고 Turn 5에 대한 응답을 한다. 직전까지 통화 종료를 시도했던 사용자였음에도 불구하고, 에이전트가 대화를 이끄는 상황에서 협조적으로 대화를 이어나가는 경향성이 보이는 것을 확인할 수 있었다.

D1 유형은 Heavy User 그룹에서는 Top 1의 순위를, Light User 그룹에서는 3위를 기록한 어려움 유형이다. 각 사용자 그룹이 이 어려움에 대응하는 방식 혹은 경향성에 차이가 있음을 위 사례들을 통해 확인할 수 있다.

4-4-2. 어려움 유형 D9에 대한 대응 방식

각 사용자 그룹별로, 특수성이 있는 몇 가지의 어려움 유형에 대해 선별적으로 Call log를 확인해 반응 혹은 대응 방식을 살펴보고자 한다. 그 중 D9 (에이전트가 하는 말을 내가 끊을 수 없음) 관련한 Call log를 탐색적으로 확인해본다. D9은 Heavy User 그룹에서만 보고된 어려움 유형이다. 일반적으로, Barge-in이 허용되지 않는 시스템 구성에서, ASR 결과는 인식이 시작된 지점부터의 부분만을 결과로 출력한다. 이 때문에, NLU가 해당 구간에 대해 잘못된 추론을 하게 될 가능성이 높다는 단점이

있다. 사용자가 VUI를 사용하면서 해당 문제를 직접적으로 맞닥뜨릴 수 있는 상황은 다음과 같다.

1. 에이전트가 사용자 발화를 무시하고, 계속 발화 진행: 사용자가 에이전트가 말하는 도중에 자신이 발화를 하게 되어도, 에이전트는 계속 발화를 진행한다. 이 과정에서, 사용자는 에이전트로부터 ‘무시받는다’라는 감정을 느낄 수 있다 (P11의 사례에서 확인).
2. 에이전트가 알아들을 수 있도록 사용자가 정확히 말했는데도, 에이전트는 못 알아들었음을 표현하는 ‘죄송한데, ...’와 같은 대화 수정 (Repair) 유도 발화를 진행: 위 사례는, 사용자 발화의 시작점과 에이전트 발화의 끝점이 겹친 상황에서 자주 발생한다. 이런 상황은 사용자로 하여금 대화 수정 (Repair) 발화를 하게 만들고, 무엇이 잘못되었는지 이해하기 어렵게 만든다.

이 문제에 대해서 Heavy User 그룹은 다음의 형태로 대응하는 것을 확인할 수 있다.

- 발화의 방식을 달리함 (발화를 길게 끌거나, 끊어서 말하거나 혹은 큰 목소리로 말함)

위 방식으로 대응하는 것을 통해, 위 어려움에 대해 사용자의 감정이 동요하였음을 유추해볼 수 있다. 관련한 대화 사례는 다음과 같다 (표 19 참조).

표 19. D9 유형에 대해 발화 방식 변경을 통한 대응

	실제 발화		Session Viewer 결과	
1	S	언제로 예약 도와드릴까요?	S	언제로 예약 도와드릴까요?
2.1	U	내일이요. (Barge-in 실패)	U	(Barge-in 실패로 기록 안됨)
2.2	U	내일...이요.	U	네일 집

3	S	죄송한데 다시 한 번 말씀해주시겠어요?	S	죄송한데 다시 한 번 말씀해주시겠어요?
4	U	내일이요!	U	데일리
5	S	혹시 더 필요하신 거 있으신가요?	S	혹시 더 필요하신 거 있으신가요?

실제로 사용자는 Turn 1이 진행되고 있는 도중에 Turn 2.1에 해당하는 사용자 발화를 완료했음에도 불구하고, 에이전트는 Turn 1을 끊지 않고 계속 진행한다. 이를 통해 사용자는 Turn 2.2의 발화를 해야만 했는데, 이때, ‘내일’이라는 표현을 길게 끄는 방식으로 발화 형태의 변경을 준다. 이와 같은 형태의 발화는 음성 인식 결과를 더 좋게 하기 위한 명확한 발화라기 보다는, 감정적 표현에 기반한 발화 형태라고 유추해볼 수 있다. 이러한 상황에서 Turn 3의 에이전트 발화는 Turn 2.2를 이해하지 못했음을 직접적으로 드러내게 되는데 (Display), Turn 4에서 사용자는 부정적 감정의 표현을 직접적으로 표현하는 것을 확인할 수 있다. 이와 같이 에이전트의 발화를 중단할 수 없는 문제는, 사용자의 발화 형태가 감정적으로 변하는 데에 영향을 미칠 수 있을 정도의 문제라고도 볼 수 있다.

4-4-3. 어려움 유형 D6에 대한 대응 방식

Light User 그룹으로부터 보고된 어려움 유형 중, D6 (가끔 지금 무슨 말을 해야할 지 모르겠는 경우가 있음)는 4순위로 보고되며, Heavy User 그룹에서의 해당 유형이 차지하는 비율보다 더 높은 것을 확인했다. 실제로, 해당 어려움을 맞닥뜨린 Light User 그룹은 다음의 형태로 반응 혹은 대응한다.

- 사용자가 에이전트의 발화 이후 일정 시간 이상 기다리는 반응을 보임
- 머뭇거림을 나타내는 대화 표지 (어..., 아...,) 등을 사용하여

인식 오류를 유발함

Light User 그룹의 사용자는 Heavy User에 비해 에이전트의 발화 이후 상대적으로 오래 기다렸다가 발화를 하는 경우가 많았다. 이는 어려움 유형 D6에 대한 사용자 Self-report에서도 동일하게 발견되었다. 다음 사례를 통해 실제 대화 속에서 해당 어려움이 어떻게 표현되고 이에 사용자는 어떻게 대응하는지 확인할 수 있다 (표 20 참조).

표 20. D6 유형에 대한 에이전트 발화를 기다리는 형태의 반응 사례

	실제 발화		Session Viewer 결과	
1	U	혹시 런치세트는 언제 할 수 있어요?	U	혹시 런치세트는 언제 할 수 있어요?
2.1	S	런치 타임은 매일 오전 11시 반부터 5시까지예요.	S	런치 타임은 매일 오전 11시 반부터 5시까지예요.
2.2		(3.1)		(3.1)
3	U	음. 혹시 통신사 할인 되나요?	U	혹시 통신사 할인 되나요
4	S	통신사가 에스케이티나 케이티이시면 할인 받으실 수 있어요.	S	통신사가 에스케이티나 케이티이시면 할인 받으실 수 있어요.

해당 발화 사례를 통해 확인할 수 있는 것은, Turn 2.1의 에이전트 발화 이후, 사용자는 Turn 2.2에서 3.1초 가량을 기다렸다가 발화를 진행한다는 점이다. 사용자는 Turn 3에서 질문을 이어나가지만, Turn 2.2의 공백을 깨기 위해 대화 표지 (음.)를 사용한 뒤 질문을 이어나간다. 에이전트와의 대화에 대한 적극적 탐색의 방식보다 수동적 반응인 기다림이 Light User가 D6의 어려움에 반응하는 한 형태였다.

또한, 앞선 사례에서 간략히 언급되었지만, D6에 대한 반응의 하나로 대화 표지의 사용이, 음성 인식에까지 영향을 주어, 대화의 흐름을 의도에

벗어나게 만든 경우도 있었다 (표 21 참조).

표 21. D6 유형에 대한 대화 표지 사용이 인식 결과에 영향을 미친 사례

	실제 발화		Session Viewer 결과	
1	U	통신사 할인은 지금 어떻게 되죠?	U	통신사 할인은 지금 어떻게 되죠
2.1	S	통신사가 에스케이티나 케이티이시면 할인 받으실 수 있어요.	S	통신사가 에스케이티나 케이티이시면 할인 받으실 수 있어요.
2.2		(2.8)		(2.8)
3	U	어... 그.	U	어린 니
4	S	키즈 메뉴 알 수 있는 홈페이지 링크 문자로 보내드릴까요?	S	키즈 메뉴 알 수 있는 홈페이지 링크 문자로 보내드릴까요?
5	U	아. 아니요. 몇 퍼센트 할인 되나요?	U	아니요 몇퍼센트 할인되나요

위 발화를 통해 볼 때, Turn 2.2에서 발생한 2.8초 가량의 침묵 (Silence) 구간에 대해 Turn 3에서 사용자는 해당 침묵에 대해 머뭇거림, 혹은 판단의 어려움을 나타내는 대화 표지 사용으로 반응하게 되었고, 이것이 잘못된 음성 인식을 유발해 Turn 4에서 잘못된 답변이 나가는 데에 영향을 주게 되었다. 결국 이는 대화의 효율 관점에서 한 개의 Turn을 낭비하게 되는 결과를 초래했고, 이후 대화 흐름에도 영향을 미칠 수 있는 여지가 남게 되었다.

지금까지, VUI 사용 시 Heavy User와 Light User가 인식하는 특징적인 어려움 유형에 대해 각 사용자 그룹이 어떻게 반응 혹은 대응하는지 살펴보았다. 각 사용자 그룹이 어려움에 대응하는 각각의 개별적인 양상보다 경향성을 파악하는 것이 중요할 것으로 생각된다. 이후 논의 주제로 연구 결과에서 확인된 정량, 정성적 분석 내용과 Call log 기반의

내용을 토대로 Heavy User와 Light User에게 최적의 사용 경험을 줄 수 있는 VUI 구축에 대해 다룬다.

5. 논 의 (Discussion)

본 장에서는 연구를 통해 도출한 정량, 정성적 분석 결과, 그리고 Call log에 기반한 대화 분석 내용을 토대로 논의 주제를 도출하고, 그에 맞게 논의를 진행한다. 논의 진행을 통해 Heavy User, Light User 그룹의 경향성을 이해할 수 있고, 이를 고려하는 VUI 구축은 어떻게 되어야 하는지 고려사항을 도출할 수 있을 것으로 기대한다.

5-1. VUI 이용 행태로 파악한 각 사용자 그룹의 경향성

본 연구를 통해 Heavy User와 Light User 간의 VUI 이용 행태 경향성을 확인할 수 있었다. 필자는 그 경향성에 대해, Heavy User 그룹은 적극적으로 VUI를 사용하며, Light User 그룹은 수동적으로 VUI를 사용한다고 표현하고자 한다. 이것은 이용 행태와 관련된 가장 상위 개념의 차이점이라고 볼 수 있다.

5-1-1. Heavy User는 적극적이며, 의사 표현이 직설적임

Heavy User는 VUI의 사용 의도를 저해하는 요소에 대해 비교적 직접적으로 자신의 감정을 표현하고 또한 능동적으로 문제 상황을 해결하는 경향이 있다. 즉, 적극적인 사용 행태를 보여주었다. 먼저 어려움 유형의 보고 개수를 보면 그 차이를 알 수 있다. Heavy User의 경우, Light User 보다 어려움 유형을 보고한 건수가 약 2배 이상 많으며, 어려움 유형의 종류도 4가지가 더 많다. 이러한 결과는 기본적으로 Heavy User가 문제 제기에 적극적이라는 증거가 될 수 있다. 동시에 이러한 문제를 제기할 수 있다는 것은, 맞닥뜨리는 어려움 유형을 세세한 층위로 파악할 수 있다는 점을 의미하기도 한다. 예를 들어, 에이전트가 사용자의 말을 알아듣지 못하는 경우에, 만약 사용자가 VUI에 대한 이해도가 높다면, 자신의 말을 인식하기 어려워하는 것인지 아니면, 인식은 되었으나 이해 혹은 처리가

안되는 것인지에 대한 미세한 차이를 파악할 수도 있다. 이는 비단, 음성 입력과 처리 관점에서 뿐만이 아니라, 본 연구에서 다룬 15가지 어려움 유형 모두와 관련이 있다고 볼 수 있다. 즉 Heavy User는 VUI 이용의 어려움에 대해 더 세세하게 정의할 수 있고, 그것을 문제 상황으로 인식할 줄 아는 적극성이 있다고 볼 수 있다.

문제 상황의 파악은 결국 해결책의 제안 혹은 임시적이지만 해당 어려움을 맞닥뜨렸을 때의 전략의 활용 가능성과도 직접적으로 연결될 수 있다. 어려움 유형 D1의 대응 사례에서 살펴보았듯이, Heavy User 그룹은 주어진 문제에 대해 적극적이면서 효율적인 대화 수정 (Repair)을 해내는 상황을 보여주었다 (P13, P18). 이러한 방식은 추가적인 문제를 만들어 낼 소지를 줄이고, 사용자로 하여금 VUI를 통제할 수 있는 가능성을 높여준다고 할 볼 수 있다. 다시 말해, 문제 상황을 파악 할 수 있다는 것은 역으로, VUI를 사용하는 방법에 대해 능숙함을 갖추고 있다는 반증이다.

하지만 흥미롭게도, 이와 같은 대응에 있어서의 능숙함과는 별개로 문제 상황을 맞닥뜨렸을 때 Heavy User 그룹의 반응은 직설적이었다. 실제로 Call log 분석 결과, D1 유형의 어려움에 직면했을 때 P07은 “근데 말투가 왜 그래요?”라는 맥락과 관계 없이 에이전트를 지적하는 형태의 감정적 발화를 내뿜었다. D1 유형에 대한 사례 외에도 Heavy User의 Call log를 살펴보면, 전반적으로 어려움 유형을 맞닥뜨렸을 때 사용자는 에이전트에 대해 비아냥대거나 (P29), 에이전트의 발화에 신경질적으로 꼬투리를 잡는 등의 모습을 보이기도 했다 (P15). 아래는, P15가 어려움에 맞닥뜨렸을 때 했던 발화의 내용 중 일부이다.

- S: 잠시만요. 11월 26일 화요일 오후 12시 세 분으로 가능하세요. 이렇게 예약해 드릴까요?
- U: 아니요. 근데 근처에 주차장 많이 있어요?
- S: 혹시 가까운 시간으로 알아봐 드릴까요?
- U: 뭘요?

- S: 오후 다섯 시 그리고 여섯 시에 가능하세요. 어떤 시간이 괜찮으신가요?
- U: 아니... 근처에 주차장 많이 있냐구요!

위 대화에서 “뭘요”라는 반응과 “아니... 근처에 주차장 많이 있냐구요!” 등의 반응은 사용자가 현 상황에서 부정적 감정을 표현하고 있음을 알 수 있는 부분이다.

일반적으로, VUI 사용 시 에이전트에 대한 이해가 부족하거나, 익숙하지 않거나 또는 기대가 낮은 경우 에이전트에 대해 감정적으로 반응을 할 것이라 생각할 수 있다. 하지만 이번 연구 결과는 이러한 예상과는 반대로, Heavy User와 Light User 그룹 간 비교에서는 Heavy User 그룹이 전반적으로 문제 상황에 직면했을 때 에이전트에 대해 부정적 감정을 표현하는 경우가 상대적으로 많았음을 확인할 수 있다. 테스트 방식이 일반적인 실험실 세팅이 아니고, 사용자에게 테스트 공간의 자유도를 제공한 상황에서, 음성 녹음과 대화 로그를 통해 감정적 반응을 확인했기 때문에 기존 실험실 세팅에서 파악하기 힘든 현실적인 반응을 얻은 것이라 생각할 수 있다. 결국 이러한 반응을 확인했다는 점은, 해당 반응이 대화 내에서 발견되었을 때 시스템 역시 해당 사항을 문제 상황으로 받아들이고, 효율적으로 풀어낼 수 있어야 함을 시사하고 있다. 즉 VUI 디자인과 시스템은 다음의 사항에 대해 어떻게 대응해야 할 것인지에 대해 고려할 수 있어야 한다.

- 고려사항 1: Heavy User의 반응 사례에서 보았듯이, 부정적 감정의 표현이 있는 경우, 시스템은 이에 대해 효과적으로 파악할 수 있어야 한다.
- 고려사항 2: 반복적 발화는 특히 Heavy User의 부정적인 감정 표현을 직접적으로 자극하는 요소이며, 이에 대한 개선은 주의 깊게 고려되어야 한다.

5-1-2. Light User는 수동적이며, 가이드를 필요로 함

Light User는 문제 상황에 대해 Heavy User 그룹에 비해 수동적인 모습을 보였다. 이들의 VUI 이용 행태를 수동적이라고 표현할 수 있는 이유는 다음과 같다. Light User 그룹이 보고한 Top 1 어려움 유형은 D3 (시스템이 내 말을 인식하는 것에 문제가 있다고 생각됨)이다. D3는 Heavy User에도 높게 나왔으나 (3위 기록), 이 어려움 유형에 대해 Light User가 인지하는 방식은 Heavy User의 어려움 유형 언급 방식에 비해 흥미롭고 특징적이라 볼 수 있다. 앞서 연구 결과 영역에 언급하였듯이, P44는 음성 인식 오류의 가능성을 자신의 발음 문제일 수도 있다고 표현한 바 있다. 다른 어려움 유형도 마찬가지로이겠지만, D3 어려움 유형은 VUI에 대한 사용자의 Input, 즉 사용자 발화가 오류를 야기할 수 있는 가능성이 상대적으로 높다고 볼 수 있다. 그리고 이러한 특성은 사용자로 하여금 VUI 오류에 대한 원인을 스스로에게 돌릴 수 있는 가능성도 함께 내포한다. 즉 VUI 이용 시 맞이하는 어려움 유형 중 인식 오류와 밀접하게 연관된 D3를 최우선으로 꼽았다는 것은, 비단 D3 뿐 아니라, 다른 어려움 유형에 대해서도 Light User는 문제의 원인을 자신에게서 찾을 가능성이 있다는 점을 시사한다.

문제의 원인을 사용자 스스로에게서 찾는 경향성은 나아가 문제 대응에 있어서의 소극적인 태도와 연결된다고 생각할 수 있다. 앞서 Call log에서도 살펴보았듯이, 문제 상황에 대해 대화 수정을 (Repair) 하는 방식에 있어서도 Light User는 에이전트에게 의존적이었다. 본인이 문제를 파악하고, 원하는 답변을 얻어내지 못해 수정 (Repair) 해나가는 과정에서도, 에이전트의 문의에 의존하는 경향을 보인다. Light User의 수정 (Repair) 유형 중 하나의 사례를 살펴보면, 문제 상황에 대한 그들의 대응 방식을 구체적으로 알 수 있다 (P34 Call log 중 발췌).

- S: 고객님 지금 예약 도와드릴까요?
- U: 어... 아뇨 제가 나중에 전화할게요.

- S: 그러면 예약 도와드릴까요?
- U: 제가 나중에 전화 드리겠습니다.
- S: 혹시 몇 분이서 오신다고 하셨죠?
- U: 13명이예요.

위와 같이 대화 수정 (Repair) 방법을 찾지 못해 전화를 끊어야 하는 상황에서도, 마지막에 시스템이 질문한 내용에 대해 답변을 한 것을 통해 Light User가 에이전트에게 의존적이라는 점을 유추해볼 수 있다. 에이전트가 현재 사용자의 요구를 해결해 낼 수 없다는 것을 정확히 파악하고 있다면, 오히려 Heavy User의 사례처럼 주도적으로 에이전트와의 대화를 활용해서 문제를 풀어나갈 것이라 예상할 수 있다. 하지만, Light User는 에이전트가 어떤 문제 상황에 있는지 이해하지 못할 가능성이 있고, 이것은 위 사례처럼 에이전트에 의존하게 되는 경향성을 만들어 낸다고 볼 수 있다.

에이전트에 대한 의존적인 성향은 대화를 어떻게 이어나갈 것인지에 대한 가이드가 필요하다는 Light User 그룹의 Self-report에서도 확인할 수 있다. 그들은 대화에서 생각하지 못한 공백이 있거나 (P42), 처음 대화를 시작할 때 (P56) 에이전트가 대화를 이끌어나가 주기를 기대했다. 이러한 구체적인 가이드가 필요하다는 점 역시, 앞서 Call log에서 보았듯이 에이전트에 의존적인 Light User 그룹의 특성과 그 궤를 같이한다.

이 밖에도, Light User가 어려움 유형에 직면했을 때 대화 표지 (음..., 아...)를 더 많이 사용하는 점도 특징적이다. 실제로 사람과 사람 간의 대화에서 발현되는 대화 선호 (Preference) 관점을 적용시켜본다면 [35], Light User가 위 대화 표지를 많이 쓰는 이유를 알 수 있다. Preference에 대한 대화 분석 영역의 학문적 해석을 보면, 사람은 자신이 기대했던 것과 다른 형태의 답변을 대화 상대방으로부터 들었을 때, 답변을 하는 데 시간이 걸리거나, 대화 표지를 사용하게 된다고 한다. 이를 VUI 이용 상황에 적용해 보면, Light User의 대화에서 발견되는 대화 표지는 사용자가

에이전트의 성능과 문제 해결 방식에 대한 이해도가 높지 않음을 나타내는 단서로 생각될 수 있다. 더 간단히 말해, 이러한 대화 표지가 많이 나오는 사용자 발화는 VUI 이용에 문제가 있다는 증거가 될 수 있다.

VUI 이용 맥락을 ‘대화’의 관점에서 바라보면, 에이전트는 대화 상대방을 이해하는 데 어려움을 느끼고 조심스럽게 대화를 진행하는 Light User와 같은 대화 참여자에 대해, 더 섬세하게 대응할 수 있어야 한다. 그들의 발화 속에서 사용자가 어떠한 상황에 처해있는지 역으로 이해하고자 노력할 필요가 있다. 이를 VUI 기술 관점으로, 동시에 대화 디자인의 관점으로 고려해보아야 한다.

- 고려사항 3: 대화가 원활히 진행되지 않고 있다는 상황을 사용자 발화로부터 또는 사용자 발화 외의 사용자 침묵 등의 비언어적인 신호를 통해 알아낼 수 있어야 한다.
- 고려사항 4: 사용자의 머뭇거림이 감지된 경우, 적절하게 디자인된 발화로 사용자를 대화의 원래 흐름으로 이끌 수 있어야 한다.

5-2. 다양한 사용자 그룹을 고려한 VUI 개선 방향성

위 논의를 통해 도출된 4가지의 고려사항과 연구 결과를 통해 파악한 어려움 유형을 이해하는 과정을 통해 상이한 사용자 그룹 모두를 위한 VUI 개선 방향성을 설정해볼 수 있다.

5-2-1. VUI 개선 우선 순위에 대한 이해의 필요성

연구 결과를 종합해보았을 때, 각 사용자 그룹이 제기한 문제를 해결하기 위한 실제적인 단계로 VUI 구성 요소의 개선 우선순위를 정해볼 수 있다. VUI를 구성하는 기술 및 디자인 요소에는 음성 인식 (ASR), 데이터 및 자연어 이해 (Data & NLU), 대화 운영 (DM), 음성 합성 (TTS)

그리고 대화 디자인 (Conversation Design)까지 다양한 영역이 있다. 앞선 연구 결과 분석에서 도출된 15가지 어려움 유형을 토대로 하여 개선이 필요한 VUI 구성 요소에 대해 정리했다. 정리된 표를 기준으로 VUI 기술 구성요소의 개선 우선 순위를 파악할 수 있다.

VUI 구성 단위	Heavy	Light
ASR	0.26	0.17
Data & NLU	0.25	0.14
DM	0.22	0.05
Conversation Design	0.21	0.09
TTS	0.07	0.02
	1.00	0.47

우선 Heavy User와 Light User 모두 공통적으로 음성 인식과 자연어 이해에 관련한 어려움 유형을 가장 많이 보고했다. 즉 사용자 발화의 인풋과 관련된 문제의 해결은 VUI에 있어 공통적으로 중요한 주제라는 것이다. 실제적인 문제 해결의 방식은 각 기술 요소 단위 별로 상이할 수 있겠지만, 전체적인 관점에서 VUI에 대한 사용자 경험을 향상시키기 위해 해당 기술 구성 요소 개선을 중심 과제로 둘 필요가 있다는 것을 보여준다.

실제로, VUI는 여러가지 기술 영역의 집합이고, 해당 기술 영역은 현재의 산업 구조 상, 또는 AI 기술 발전 상황 상 독립적으로 고도화가 필요한 영역이다. 하지만, VUI의 발전을 위해서는 기술 개별 단위가 아니라, VUI 단위로서의 발전에 대해서도 신경을 써야 한다. 사용자가 경험하는 어려움 유형이 하나의 기술 구성 요소에만 의존적이지 않다는 것을 연구 결과를 통해서 살펴볼 수 있었다. 사용자가 맞이하게 되는, 겉으로 드러나는 현상은 각 구성 요소의 복합적인 조합으로 발현되는 것이다. 그렇기 때문에 Heavy User 그룹과 Light User 그룹의 사용자 별 개선에 앞서 VUI 발전에 대한 통섭적인 시각을 갖는 것이 실제로 중요하다. 통섭의 시각을 가질 수 있을 때 비로소, VUI 관점에서의 개선 방향성에 대해 각 기술 구성 요소,

구체적으로 해당 담당자들이 원활히 커뮤니케이션을 할 수 있으며, 사용자 중심의 더 나은 경험을 만들 수 있는 기반이 될 수 있다.

5-2-2. 소통 실패를 해결하기 위한 실질적인 방안

VUI의 개선 우선 순위를 알아내는 논의는, 통합적인 관점에서 VUI를 개선하는데 실제적인 도움을 줄 수 있다. 그러나, 본 연구에서 도출한 4가지 고려사항에 대해 논의하기 위해서는 더 실질적이고 구체적인 방안이 필요하다.

5-2-2-1. 음성 정보를 활용한 사용자 상태 이해

본 연구원은 VUI와 소통 실패가 발생한 상황에 사용자는 말의 내용이라고 볼 수 있는 ‘텍스트’ 뿐 아니라, 음성 정보로도 자신의 상태를 표현하고 있음을 확인할 수 있었다. 먼저 고려사항 1과 관련해서 음성 정보를 활용한 감정에 대한 이해의 필요성을 확인할 수 있었다. 먼저, Heavy User의 감정 표출과 관련해서는 음성의 세부 특징, 즉 높낮이 (Pitch)나 음량 (Volume) 또는 세기 (Energy) 등을 활용해 사용자의 상태를 파악할 필요성이 있다. 현재의 음성 인식 분야에서는 이와 같이 음성 정보를 활용해 사용자의 감정을 인식할 수 있도록 진행하는 기술적 연구가 활발히 진행되고 있다 [2, 15]. 오늘날의 VUI는 이와 같이 음성 정보를 활용한 사용자 감정을 이해하기 위해 관련된 기술 도입이 필요한 시점이라는 것을 확인할 수 있다.

이와 동시에, 기술적인 난이도는 높지 않지만, 유효하게 사용할 수 있는 음성 정보로 침묵을 들 수 있다. 세 번째 고려사항을 통해 도출하여 보았듯이, Light User 그룹이 어려움에 처했을 때 발현되는 주된 현상 중 하나인 침묵은 주의 깊게 다뤄져야 할 신호 중 하나이다. 음성 인터페이스에서 침묵이 있는 구간을 잡아내는 것은 비교적 어렵지 않다. 이렇게 사용자가 현재 상태에서 이상 상황에 있다는 것을 음성 정보로

알아낸다면, 본 연구에서 도출한 다른 두 가지 고려사항 (고려사항 2, 4)을 해결하기 위한 논의를 진행시킬 수 있다.

5-2-2-2. 소통 실패에 대비한 대화 디자인의 필요성

상이한 사용자 그룹이 처한 어려움, 즉 소통 실패 영역에 대한 복귀 전략이 더 원활하게 진행될 수 있도록 VUI는 사용자를 가이드할 수 있어야 한다. 앞서 논의했듯이 챗 인터페이스는 이러한 복귀 전략에 대한 가이드로 GUI의 장점을 활용한다. 음성 중심의 VUI는 앞서 언급한 음성 정보를 활용해 사용자의 대화 수정 과정 (Repair Practice)을 도울 필요가 있다.

수정 과정에 대한 대비 혹은 가이드는 상당 부분 대화 디자인과 그를 운영하는 대화 운영 (DM) 영역에서 관장해야 한다. 고려사항 2와 관련하여, 소통 실패에 의해 반복적인 발화가 발생하는 경우에 대비한 대화 정책이 필요하다. 앞서 살펴 보았듯, 반복적인 발화는 Heavy User에게 있어 가장 문제되는 어려움 유형이었다. 필수적으로 반복적 안내가 필요한 대화에서도 부정적 감정 표현 등이 감지되었을 때는 이에 대응할 수 있는 정책이 필요하다.

부정적 감정 표현이 발생하는 구간은 대화에 있어 최종적인 소통 실패가 발생한 구간으로 간주해볼 수 있다. 다시 말해, 사용자가 대화를 수정하는 과정을 더 이상 진행하지 않겠다고 표현한 지점과 다름 없다고 볼 수 있다. 이런 경우, AiCall과 같은 목적 지향형 대화 시스템은 대화 진행을 포기할 수 있어야 한다. 대화 디자인 상, 에이전트의 관점에서 더 많은 대화 수정 과정들이 준비되어 있다고 하더라도, 사용자의 부정적 감정 표현을 감지하게 되었다면, 대화 수정 과정은 중단될 수 있어야 한다. 예를 들면, 다음과 같은 대화 진행이 필요할 수 있다.

표 22. P15 (Heavy User)의 부정적 감정에 대한 대응 방안 재구성

Turn	Speaker	Utterance	비고
1	S	잠시만요. 11월 26일 화요일 오후 12시 세 분으로 가능하세요. 이렇게 예약해 드릴까요?	
2	U	아니요. 근데 근처에 주차장 많이 있어요?	
3	S	혹시 가까운 시간으로 알아봐 드릴까요?	
4	U	뭘요?	
5	S	오후 다섯 시 그리고 여섯 시에 가능하세요. 어떤 시간이 괜찮으신가요?	
6	U	아니... 근처에 주차장 많이 있냐구요!	사용자의 부정적 감정 표출
7	S	제가 말씀하신 내용을 알아듣지 못해서 죄송합니다. 요청하신 내용 해결하실 수 있게, 매장 직원 분 연결해드려도 괜찮을까요?	음성 정보를 반영한 대응 예시

위 재구성된 예시에서, 에이전트는 사용자의 부정적 감정 표현을 인지하고 (Turn 6), 대화 진행이 더 이상 어렵다는 사실을 안내한다 (Turn 7). 이와 같은 대응은 곧 사용자의 목적 성취에 초점을 둔 대응이며, 대화 시스템은 한계를 파악하고 그 이상의 시도를 하지 않아야 한다는 점에서 시사하는 바가 있다.

Light User의 소통 실패를 암시하는 침묵과 관련된 대응 역시 VUI에서 중요하다. 구체적으로 말하면 침묵이라는 비언어적 신호는 연속적인 대화가 일어날 것으로 가정된 VUI에서 특히 중요하다. AiCall과 같은 전화 상의 VUI가 그 예시가 될 수 있으며, 그 외에도 Google의 Continued

Conversation과 같이 기능 상 연속 대화를 제공하는 VUI에서 침묵은 중요한 의미를 가질 수 있다.

일반적으로 사용자의 침묵이 감지되면, 에이전트는 사용자가 대화에 참여하고 있는지 여부를 확인한다. AiCall 역시 사용자가 대화에 참여하고 있는지에 대한 여부를 문의하는 방식으로 대화가 디자인되어 있다. 하지만, Light User의 어려움 유형을 확인했을 때, 침묵은 대화에 참여하고 있는 중에도 발생할 수 있다. 그러므로 대화를 디자인하고, 정책을 정의할 때 사용자 침묵에 대해 더 세부적으로 정의해낼 수 있어야 한다. 즉, 사용자가 대화를 이탈한 상황인지 그렇지 않고 다음 발화를 말하기 위한 준비 과정인지 구분해내는 것이 중요하다.

그런데, 한 번의 Turn-taking 관점에서 침묵의 의미에 대해 세부적으로 정의하기는 어렵다. 왜냐하면 대화가 진행되면서 침묵의 횟수가 빈번히 발생하는 것이 대화 진행에 있어 일련의 소통 실패가 반복적으로 나타난다는 증거가 될 수 있기 때문이다. 그러므로 사용자 침묵을 대화 디자인으로 정의할 때는 일회성의 사용자 침묵에 대한 대화 정책 정의에 더해, 반복적인 사용자 침묵이 대화 중에 나타날 때를 대비한 대화 디자인도 필요하다.

반복적인 사용자 침묵이 대화 중에 나타난다면, 이는 소통 실패의 근거로 볼 수 있고, 연구 결과의 Light User의 어려움 유형을 통해 확인한 것처럼 사용자가 가이드를 필요로 하는 상황으로 유추해볼 수 있다. 이러한 경우를 대비해 반복적인 사용자 침묵에 대한 다음과 같은 가이드 발화 역시 준비해 놓을 필요가 있다.

표 23. 사용자 침묵에 대한 상세 정의와 응대 예시

Case	Goal	System Utterance (AiCall 예시)
------	------	---------------------------------

일회성 사용자 침묵	사용자의 대화 가능 여부를 확인함	고객님, 통화 가능하신가요?
반복적 사용자 침묵	사용자에게 대화를 원활히 이어갈 수 있는 안내성 발화를 제공	1. 예약 진행 흐름의 경우: 고객님, 질문 다시 한 번 드릴까요? 2. 단순 질문 흐름의 경우: 고객님 예약이나 매장 이용 전반에 대해 다른 문의사항 있으신가요?

위 표에서와 같이 사용자 침묵에 대해 대응하는 방식도 크게 일회성, 그리고 반복적 사용자 침묵을 나누어 디자인 할 수 있고, 반복적 사용자 침묵의 상황에서도 목적 지향형 대화의 경우와 일반적인 질문과 답변 경우를 나누어 대화를 디자인할 수 있다.

이러한 대응 방식은, Light User 그룹의 사용자처럼 VUI 이용의 소통 실패 단계에서 적극적인 문제 해결을 하기 어려운 사용자 그룹에 대해 도움이 되는 방식이라 생각할 수 있다.

위 논의 내용을 통해 Heavy User와 Light User 그룹이 본질적으로 어려움을 겪는 문제를 분석하고, VUI 디자인 시 고려해야 할 사항에 대해 알아보았다. 구체적으로 도출된 4가지의 고려 사항을 해결하기 위해, 즉 소통 실패 상황에서 사용자의 명확하거나 안정적인 대화 복귀가 가능하도록 돕기 위해 VUI가 구성되어야 방향성을 알아보았다. 해당 논의 방향성은 앞으로 상이한 사용자 그룹을 수용하기 위해 VUI가 발전해 나가야 하는 방향성을 제시한다.

6. 연구의 한계 (Limitations)

본 연구는 AiCall이라는 전화 상의 VUI를 활용하여 Heavy User와 Light User/간 이용 행태의 차이를 알아보았다. 각 사용자 그룹은 어떤 형태의 어려움을 겪는지 대화의 소통 실패 관점으로 알아보았다. 이를 근거로 하여 사용자가 소통 실패로부터 명확하게, 그리고 안정적으로 대화를 이어나갈 수 있도록 복귀 가이드를 제공하는 방안 역시 다루었다. 그럼에도 불구하고, 본 연구는 한계점 역시 존재한다. 첫째로, 연구의 진행 시점은 2019년 11월로, AiCall이 아웃백 스테이크하우스 미금점이나, 홍보 페이지를 통해 실사용자에게 선보이기 전이다. 정식 릴리즈 전의 제품을 활용한 연구였기 때문에, 테스트 대상자들은 사내 직원들이었다. 물론 Naver라는 회사가 기본적으로 IT회사임을 감안할 때, 사내 직원들을 대상으로 한 연구의 결과는 실제 사용자에게서 얻을 수 있는 결과와 다를 가능성이 있다. 2020년 상반기, AiCall이 출시된 이후 무작위로 선정된 사용자 테스트를 통해, 본 연구의 결과와 실제 사용자 대상 테스트 결과를 비교해 나갈 예정이다.

또한, 본 연구는 Heavy User 그룹과 Light User 그룹을 나눈 뒤 이들의 초기 사용 행태에 집중하여 연구 결과 도출 및 논의를 진행하였다. 하지만, 실제로 Heavy User 그룹과 Light User 그룹 각각 지속적으로 VUI를 이용하게 되는 경우 각 사용자 그룹이 어떤 형태로 변화해 나갈 것인지에 대한 관찰 및 이해는 이번 연구에서 진행되지 않았다. 앞으로의 연구에서는 동의 및 승낙을 받은 사용자에게 한해, 본 연구의 사용자 그룹을 나누는 기준을 적용해, 각각의 사용자 그룹이 반복적으로 에이전트를 이용하게 되면서 어떤 변화를 겪게 되는지 알아볼 예정이다.

마지막으로, 본 연구가 전화 상의 VUI를 대상으로 진행되었기 때문에, VUI 활용 기기가 달라지는 경우에는 다른 형태의 어려움 유형이 도출될 수 있으며 이에 따른 고려사항이 발생할 수 있다. 예를 들어, 전화 상의 VUI는

필연적으로 개인 사용을 기반으로 한 것이고, 호출어가 존재하지 않으며, 그렇기 때문에 더욱 대화의 흐름이 중요해질 수 있다. 물론 본 연구의 결과와 논의 주제가 VUI에 대해 일반적으로 적용될 수 있는 부분이 존재하지만 스마트 스피커나 스마트 폰의 VUI의 경우 그 특수성이 달라질 수 있으므로, 해당 환경에 대한 독립적인 연구가 필요하다.

7. 결 론 (Conclusion)

본 연구는 Heavy User와 Light User 그룹을 기술 수용도 관점으로 나눠보고, 상이한 사용자 그룹을 만족시킬 수 있는 VUI가 되기 위해서는 어떤 고려사항이 필요한지 탐구하였다. 구체적으로, 먼저 사용자 그룹이 VUI를 사용하면서 겪는 어려움은 무엇인지를 알아보고 유형화한 뒤 주요 어려움 유형에 대해 사용자 그룹 별로 어떻게 반응 혹은 대응하는지 알아보았다. 이후 해당 연구 결과를 토대로 상이한 사용자 그룹을 위한 맞춤형 VUI가 되기 위한 고려사항과 이를 실제적으로 해결하기 위한 VUI 기술, 디자인 관점의 방법에 대해 논의하였다.

이 연구는 VUI 활용 제품과 서비스가 늘어나는 현재의 시점에, VUI도 이제는 다양한 사용자 층의 사용 경험을 고려할 필요가 있음을 연구 결과와 논의를 통해 제안하였다. 구체적으로 본 연구가 HCI 커뮤니티와 VUI 관련 산업계에 기여할 수 있는 바는 다음과 같다.

- 기술 수용도 관점에 따라 VUI가 초기 다수자 (Early Majority)에게까지 받아들여지고 있는 현 시점에서, 출시 초기부터 다양한 사용자 층을 받아들일 가능성이 있음을 염두에 두고 VUI 관련 제품 혹은 서비스를 개발해야 한다
- HCI 커뮤니티는 VUI에 대한 특정 사용자 층 연구 흐름을 확장해, 하나의 VUI에 대해서도 서로 다른 특징을 가진 사용자 층이 사용하고 경험하는 바가 다를 수 있음을 이해해야 한다
- Heavy User와 Light User가 VUI에 대해 경험하는 어려움이 서로 다를 수 있음을 이해하고, 그것을 VUI 기술 관점에서 그리고 대화 디자인 관점에서 어떻게 풀어나갈 수 있을지 고민해야 한다
- 특히 VUI 기술 관점에서는 텍스트를 기반으로 한 언어 정보 외에 음성 정보도 대화 시스템 전반에서 활용할 수 있는

방안을 확보해야 하며, 이것이 서로 다른 사용자 그룹의 특징을 이해하기 위한 출발점임을 인지하고 있어야 한다

- 대화 디자인 관점에서는 사용자 발화에 대한 다양한 층위의 정보를 활용하게 될 상황을 고려해 전체 대화 공간을 구성할 때의 복잡성을 어떻게 다루어야 할 것인지 고민해야 한다

참고 문헌

- [1] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: repair strategy preferences for conversational breakdowns. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1 – 12.
- [2] Abdul Malik Badshah, Jamil Ahmad, Nasir Rahim, and Sung Wook Baik. 2017. Speech emotion recognition from spectrograms with deep convolutional neural network. In 2017 international conference on platform technology and service (PlatCon). IEEE, 1 – 5.
- [3] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1 – 13.
- [4] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? voice-driven interfaces and young children's communication repair strategies. In Proceedings of the 17th ACM Conference on Interaction Design and Children. 337 – 348.
- [5] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers* 31, 4 (2019), 349 – 371.
- [6] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad,

Cosmin Munteanu, et al. 2019. What makes a good conversation? challenges in designing truly conversational agents. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1 – 12.

[7] Michael H Cohen, Michael Harris Cohen, James P Giangola, and Jennifer Balogh. 2004. Voice user interface design. Addison–Wesley Professional.

[8] Juliet Corbin and Anselm Strauss. 2014. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.

[9] Jacob Devlin, Ming–Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre–training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).

[10] Anne Drabble. 2015. Characteristics of Heavy Users and Light Users’ Contributions Using a Social Networking System to Enhance Literacy Learning. *Ubiquitous Learning: An International Journal* 8, 3 (2015).

[11] Lila J Finney Rutten, Erik M Augustson, Kelly A Doran, Richard P Moser, and Bradford W Hesse. 2009. Health information seeking and media exposure among smokers: a comparison of light and intermittent tobacco users with heavy users. *Nicotine & Tobacco Research* 11, 2 (2009), 190 – 196.

[12] Asbjørn Følstad and Cameron Taylor. 2019. Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives. In *International Workshop on Chatbot Research and Design*. Springer, 201 – 214.

[13] Ronald E Goldsmith. 2000. Characteristics of the heavy user of fashionable clothing. *Journal of Marketing Theory and Practice* 8, 4 (2000), 21 – 28.

[14] Florian Habler, Marco Peisker, and Niels Henze. 2019. Differences between smart speakers and graphical user interfaces for music search considering gender effects. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. 1 – 7.

[15] Kun Han, Dong Yu, and Ivan Tashev. 2014. Speech emotion recognition using deep neural network and extreme learning machine. In *Fifteenth annual conference of the international speech communication association*.

[16] Randy Allen Harris. 2004. *Voice interaction design: crafting the new conversational speech systems*. Elsevier.

[17] June Kaminski. 2011. Diffusion of innovation theory. *Canadian Journal of Nursing Informatics* 6, 2 (2011), 1 – 6.

[18] Joel D Killen, Stephen P Fortmann, Michael J Telch, and Barbara Newman. 1988. Are heavy smokers different from light smokers?: A comparison after 48 hours without cigarettes. *Jama* 260, 11 (1988), 1581 – 1585.

[19] A Baki Kocabalil, Liliana Laranjo, and Enrico Coiera. 2018. Measuring user experience in conversational interfaces: a comparison of six questionnaires. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference* 32. 1 – 12.

[20] Younghwa Lee, Kenneth A Kozar, and Kai RT Larsen. 2003. The technology acceptance model: Past, present, and future. *Communications of the Association for information systems* 12, 1 (2003), 50.

[21] James R Lewis. 2016. Standardized questionnaires for voice interaction design. *Voice Interaction Design* 1, 1 (2016).

[22] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 265 – 268.

[23] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286 – 5297.

[24] Jianliang Meng, Junwei Zhang, and Haoquan Zhao. 2012. Overview of the speech recognition technology. In *2012 fourth international conference on computational and information sciences*. IEEE, 199 – 202.

[25] Robert J Moore, Raphael Arar, Guang-Jie Ren, and Margaret H Szymanski. 2017. Conversational UX design. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 492 – 497.

[26] Christine Murad, Cosmin Munteanu, Benjamin R Cowan, and Leigh Clark. 2019. Revolution or Evolution? Speech Interaction and HCI Design Guidelines. *IEEE Pervasive Computing* 18, 2 (2019), 33 – 45.

[27] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 1 – 7.

[28] Larry P Pleshko and Sarah Al-Houti. 2012. Heavy versus light users: A preliminary study of behavior patterns in retail services. *Academy of Marketing Studies Journal* 16 (2012), 61.

[29] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In proceedings of the 2018 CHI conference on human factors in computing systems. 1 – 12.

[30] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. " Alexa is my new BFF" Social Roles, User Satisfaction, and Personification of the Amazon Echo. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2853 – 2859.

[31] Kwansub Shim, Seoul Seodaemun-gu, Carl Changseok Yoon, and Jungwoo Lee. 2015. Comparative Study of Heavy and Light Users of Popular Smart Apps. In Proceedings of The Fourth International Conference on Informatics & Applications. 260 – 267.

[32] Eunwoo Song, Kyunguen Byun, and Hong-Goo Kang. 2019. ExcitNet vocoder: A neural excitation model for parametric speech synthesis systems. In 2019 27th European Signal Processing Conference (EUSIPCO). IEEE, 1 – 5.

[33] Norbayah Mohd Suki and Norazah Mohd Suki. 2007. Mobile phone

usage for m-learning: comparing heavy and light mobile phone users. Campus-Wide Information Systems (2007).

[34] Mikael L Tedjopranoto, Andreas Wijaya, Levi Hanny Santoso, and Derwin Suhartono. 2019. Correcting Typographical Error and Understanding User Intention in Chatbot by Combining N-Gram and Machine Learning Using Schema Matching Technique. *International Journal of Machine Learning and Computing* 9, 4 (2019), 471 – 476.

[35] Paul Ten Have. 2007. *Doing conversation analysis*. Sage.

[36] Hsu-Ju Teng, Jia-Jen Ni, and Hsiao-Han Chen. 2018. Relationship between e-servicescape and purchase intention among heavy and light internet users. *Internet Research* (2018).

[37] Helen Tibboel, Jan De Houwer, Adriaan Spruyt, Damien Brevers, Emmanuel Roy, and Xavier Noël. 2015. Heavy social drinkers score higher on implicit wanting and liking for alcohol than alcohol-dependent patients and light social drinkers. *Journal of behavior therapy and experimental psychiatry* 48 (2015), 185 – 191.

[38] Hannu Verkasalo, Carolina López-Nicolás, Francisco J Molina-Castillo, and Harry Bouwman. 2010. Analysis of users and non-users of smartphone applications. *Telematics and Informatics* 27, 3 (2010), 242 – 255.

[39] Reinout W Wiers, Nieske Van Woerden, Fren T Y Smulders, and Peter J De Jong. 2002. Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of abnormal psychology* 111, 4 (2002), 648.

[40] J Wilke, F McInnes, Mervyn A Jack, and P Littlewood. 2007. Hidden menu options in automated human – computer telephone dialogues: dissonance in the user’s mental model. *Behaviour & Information Technology* 26, 6 (2007), 517 – 534.

[41] Maria Wolters, Kallirroi Georgila, Johanna D Moore, Robert H Logie, Sarah E MacPherson, and Matthew Watson. 2009. Reducing working memory load in spoken dialogue systems. *Interacting with Computers* 21, 4 (2009), 276 – 287.

[42] Xiangyang Zhou, Lu Li, Daxiang Dong, Yi Liu, Ying Chen, Wayne Xin Zhao, Dianhai Yu, and Hua Wu. 2018. Multi–turn response selection for chatbots with deep attention matching network. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1118 – 1127.

Abstract

Understanding Differences between Heavy Users and Light Users in Difficulties with Voice User Interfaces

Hyunhoon Jung

Digital Contents and Information Studies,
The Graduate School of Convergence Science and
Technology,
Seoul National University

Voice User Interfaces (VUIs) are growing in popularity. Recent market research has shown that VUIs are being adopted by the ‘early majority’ based on the ‘Diffusion of Innovation’ theory. This means that VUIs are not just for the ‘innovators’ or the ‘early adopters.’ While it is important to understand the usage patterns of different user groups, few studies focus solely on these differences. In this paper, the VUIs user group are divided into two categories, heavy users and light users, in terms of the frequency and recency of use. To understand the differences between heavy users and light users in difficulties, I conduct user study using the restaurant reservation VUI, AiCall, which I have participated in designing and development. From the result of the user study, I found out that each user group had a different way of dealing with the difficulties; heavy users were active and aggressive, while light users were passive. I hope that the findings and the discussions of this study would be

used to design and develop VUIs for both user groups.

Keywords : Voice User Interface (VUI), Conversational User Interface (CUI),
Voice UX, Conversational UX

Student Number : 2017-29610